REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
29-08-2013	Final Report	7-Jun-2007 - 6-Jul-2013

4. TITLE AND SUBTITLE

Time of Flight Estimation in the Presence of Outliers: A biosonar-inspired machine learning approach

5a. CONTRACT NUMBER
W911NF-07-1-0256

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Nathan Intrator, Leon N Cooper

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Brown University
Office of Sponsored Projects
Box 1929
Providence, RI 02912 -9093

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
52890-LS.10

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
When the Signal-to-Noise Ratio (SNR) falls below a certain level, the error of the Time-of-Flight (ToF) Maximum Likelihood Estimator (MLE) increases abruptly due to the well known threshold effect. Nevertheless, operating near and below the threshold SNR value might be necessary for many remote sensing applications due to power-related constraints. These constrains may include a limit on the maximum power of a single source pulse or a limit on the total power used by multiple signals transmitted during a single measurement. For narrowband

15. SUBJECT TERMS
sonar, underground installations, biosonar, remote sensing, sonar resolution, sonar accuracy, sonar energy consumption

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Leon Cooper
UU	UU	UU	UU		19b. TELEPHONE NUMBER: 401-863-2172

Report Title

Time of Flight Estimation in the Presence of Outliers: A biosonar-inspired machine learning approach

ABSTRACT

When the Signal-to-Noise Ratio (SNR) falls below a certain level, the error of the Time-of-Flight (ToF) Maximum Likelihood Estimator (MLE) increases abruptly due to the well known threshold effect. Nevertheless, operating near and below the threshold SNR value might be necessary for many remote sensing applications due to power-related constraints. These constrains may include a limit on the maximum power of a single source pulse or a limit on the total power used by multiple signals transmitted during a single measurement. For narrowband signals, the threshold effect emerges mostly due to outliers induced by local maxima of the autocorrelation function of a source signal. Following the previously explored path of biosonar-inspired echo processing, in this research we introduce new methods for ToF estimation in the presence of outliers. The proposed methods employ a bank of phase-shifted unmatched filters for generating multiple biased but only partially correlated estimators (multiple experts). Using machine-learning techniques, the information from the multiple experts is combined together for improving the near-the-threshold ToF estimation from a single echo. We describe methods for ToF estimation from single and multiple pulses as well as the method for improving the energy efficiency of the estimation.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received		Paper
08/29/2013	9.00	Alexander Apartsin, Nathan Intrator, Leon Cooper. Time-of-Flight Estimation in the Presence of Outliers Part II-Multiple Echo Processing, Geoscience and Remote Sensing, IEEE Transactions on, (07 2013): 0. doi:
11/02/2012	8.00	Alexander Apartsin, Leon N Cooper, Nathan Intrator. Semi-coherent time of arrival estimation using regression, Journal of the Acoustical Society of America, (08 2012): 832. doi:
TOTAL:	**2**	

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received		Paper
TOTAL:		

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received	Paper	
02/08/2010	1.00	K. Kim, N. Neretti, N. Intrator. MAP Fusion Method for Super-resolution of Images with Locally Varying Pixel Quality, (01 2008)
02/16/2010	3.00	A. Apartsin, N. Intrator, L. Cooper. Time of arrival estimation in low SNR, (02 2009)
02/16/2010	2.00	N. Intrator, L. Cooper. Detection of underground installations in hostile environments a biosonar application: Research report, (12 2009)
08/31/2010	5.00	Sasha Apartsin, Leon N. Cooper, Nathan Intrator. BIOSONAR-INSPIRED SOURCE LOCALIZATION IN LOW SNR, (08 2010)
09/15/2011	6.00	Leon N Cooper, Sasha Apartsin, Nathan Intrator. SEMI-COHERENT TIME-OF-ARRIVAL ESTIMATION REVISITED: MACHINE LEARNING BEATS THE MATCHED FILTER, Journal of the Acoustical Society of America (09 2011)
11/01/2012	7.00	Alexander Apartsin, Leon N Cooper, Nathan Intrator. Time of Flight Estimation in the Presence of Outliers Part I-Singe Echo Processing, ( )
TOTAL:	**6**	

Number of Manuscripts:

Books

Received	Paper
TOTAL:	

Patents Submitted

Patents Awarded

Awards

Leon N Cooper, Susan Culver Rosenberger Medal, Brown University, 2013

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Alexander Apartsin	0.75	
FTE Equivalent:	**0.75**	
Total Number:	**1**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Leon N Cooper	0.10	Yes
Nathan Intrator	0.10	
FTE Equivalent:	**0.20**	
Total Number:	**2**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	0.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:.....	0.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:.....	0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	0.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME
Alexander Apartsin
Total Number: 1

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

See attached.

Technology Transfer

Time of Flight Estimation in the Presence of Outliers

A biosonar-inspired machine learning approach

Leon N Cooper, Nathan Intrator

Final Research Report

August 2013

***Abstract*—** When the Signal-to-Noise Ratio (SNR) falls below a certain level, the error of the Time-of-Flight (ToF) Maximum Likelihood Estimator (MLE) increases abruptly due to the well-known threshold effect. Nevertheless, operating near and below the threshold SNR value might be necessary for many remote sensing applications due to power-related constraints. These constraints may include a limit on the maximum power of a single source pulse or a limit on the total power used by multiple signals transmitted during a single measurement. For instance, these requirements emerge in military applications that require low powered pulses for the measurements process to stay undetected by the adversary or a limit on the total energy used if measurements are performed by a mobile robot equipped with an autonomous power source (battery).

For narrowband signals, the threshold effect emerges mostly due to outliers induced by local maxima of the autocorrelation function of a source signal. Following the previously explored path of biosonar-inspired echo processing, in this research we introduce new methods for ToF estimation in the presence of outliers. The proposed methods employ a bank of phase-shifted unmatched filters for generating multiple biased but only partially correlated estimators (multiple experts). Using machine-learning techniques, the information from the multiple experts is combined together for improving the near-the-threshold ToF estimation from a single echo. We describe methods for ToF estimation from single and multiple pulses as well as the method for improving the energy efficiency of the estimation.

I. INTRODUCTION

The classical Time of Flight (ToF) estimation method employs a matched filter [20] for the estimation of the two-way travel time. Since the returned signal is usually corrupted by an additive noise, the resulting estimator produces a value which is only an approximation of the true two-way travel time. The average magnitude of the squared estimation error depends on the amount of additive noise and a shape of the source waveform [23]. For moderate levels of noise and for source pulses with sufficiently wide bandwidth, the Mean Square Error (MSE) of the classical estimator closely follows the Cramer-Rao Low Bound (CRLB) which is the best achievable performance for any "good" estimator [20]. In the theory of optimal receivers, these conditions correspond to a receiver being in the coherent state [11].

However, in many practical cases, active sonar or other remote sensing devices are employed under conditions that no longer ensure a coherent reception and estimation. An application might impose power and bandwidth constraints on the source pulses generated by a transmitter. For instance, the maximum power of an individual pulse could be restricted by the limitations of electronic or acoustic equipment employed at the transmitter. For military applications, it is sometimes important to keep the source power low to avoid detection by the adversary. The low accuracy of individual estimations could be somehow compensated for by averaging (fusing) multiple independent measurements. However, the usage of multiple source pulses for a single estimation might be limited by constraints imposed on the total measurement time (or, equivalently, on a number of measurements) and by constrains on the total employed power. The later is crucial for measurements that are done by a mobile robotic sensor with limited autonomous power source (a battery) [26]. Moreover, there also might be constraints on a shape of the transmitted source signal. For instance, underground exploration by low–powered seismic pulses employs low-bandwidth source waveforms since the high frequency harmonics are attenuated rapidly in the ground [27].

The classical approach based on the application of matched filter and on a simple averaging of multiple measurements is not adequate for these constrained cases. It has been shown that if the level of noise rises above a certain level, the mean squared error of the conventional ToF estimator abruptly diverges from CRLB producing the well-known threshold effect [23]. Moreover, the threshold effect intensifies when a source pulse waveform has low bandwidth spectrum as is sometimes the case in underground explorations. The classical approach does not provide an efficient way for taming the Mean Square Error (MSE) of near-the-threshold ToF estimation and it does not address the energy-related constraints in a proper way. Therefore, the alternative approach is required for improving the efficiency of ToF estimation that is carried out in the presence of the threshold effect and within the imposed power limitations.

It appears that some of the ideas for improving the accuracy of near-the-threshold ToF estimation could be borrowed from nature. Many echolocating animals use mechanical waves (e.g. airborne ultrasound or underground infrasound) for communications and navigations. Naturally, biological systems have many constraints similar to those described above and yet, the echolocating animals demonstrate striking ability

for efficient extraction of information from very noisy signals [9]. Moreover, a biosonar processing system employs individual elements (e.g. cells) with insufficient time-frequency sensitivity for reliably estimating the signal parameters directly. However, organized in massively parallel sensory and computational networks, these elements can function together with astonishing accuracy [16].

A biologically inspired approach to the ToF estimation tries to extract additional information from the received signal by employing a bank of filters instead of a single matched filter as in the classical theory. The partial information obtained by the application of individual filters is combined together using robust fusion methods in order to produce an estimator that outperforms the classical matched filter.

There is a significant body of research focusing on analyzing and mimicking biosonar systems. In [12], a filter bank of 22 combined band pass and low pass filters was inspired by the echolocating characteristics of *eptesicus fuscus* (the big brown bat). The information from these filters was combined using three fusion techniques including average-like and voting-like methods. This work has shown that in high noise settings, accuracy comparable to that of the classical approach could be achieved by intelligently fusing information from multiple bandwidth-constrained filters. In [17], the two-glint resolution of constrained sonar was analyzed using similar approach that included a filter bank of 81 low-bandwidth filters followed by a form of template matching. The resolution achieved by this approach was far beyond the resolution of individual processing elements (filters).

In this work, we continue with biosonar-inspired approach by analyzing the threshold effect for simple narrowband source signals. Identification of the major source of the rapid degradation of accuracy caused by the threshold effect motivates the design of a family of full bandwidth phase-shifted filters. These filters combined together provide more information compared to the matched filter alone. This extra information can be used for mitigating the threshold effect and, thus, for improvement in the accuracy of a single measurement. We employ machine learning techniques for extracting the information from the vector of responses generated by the application of these multiple filters. For this purpose, we construct a classifier to assign a label to a measurement based on the vector of filters' responses computed on the retuned signal. Using the assigned label, the classical estimate (obtained using the matched filter) is corrected to account for the expected bias that appears due to the threshold effect. Moreover, the information supplied by the classifier is used for fusing estimates from multiple pulses into a robust single estimate that weights individual measurements according to the estimated degree of uncertainty. Finally, we describe the method which utilizes the classifier for adaptively controlling the number of pulses required for achieving the desired accuracy.

II. TIME-OF-FLIGHT PROBLEM AND THE THRESHOLD EFFECT

In this section, we provide an analytical treatment of the threshold effect associated with Time of Flight (ToF) Maximum Likelihood Estimator. We derive an approximation for the probability of an outlier event using fundamental properties of narrowband source signal, namely its centralized bandwidth to central frequency ratio. In the next section, we will extend the analysis for biased estimators and show that it is possible to devise a family of biased estimators that are not completely correlated.

The Time of Flight estimation (ToF) problem also known as "Time of Arrival" (ToA) estimation problem arises in the context of Radar, Sonar and other remote sensing applications. In discrete formulation, the problem could be stated as follows [20]. An acoustic source emits a signal waveform $s(t)$ and, then, the reflected signal contaminated by Additive White Gaussian Noise (AWGN) $n(t)$ is picked up by a receiver after a time delay t_0 . Then, the received signal is given by:

$$r(t) = s(t - t_0) + n(t)$$

The goal of ToF estimation is to recover a value of unknown time delay t_0 based on the received signal $r(t)$ and on the known source waveform $s(t)$. We usually seek an estimator $\hat{t}_0(r)$ that minimizes the MSE cost function. Since noise samples $n(t)$ are i.i.d. random variables with a probability density $\boldsymbol{N}(0,\sigma^2)$, the conditional probability density function for received signal given the time delay is

$$\Pr_{r|t}(r(t) \mid \tau) = \sqrt{2\pi\sigma^2}^{-N} \exp\left(-\frac{1}{2\sigma^2} \sum_t \left(r(t) - s(t-\tau) \right)^2 \right)$$

where N is the number of waveform samples recorded during an observation interval. Assuming no prior knowledge of values of time delay (uniform distribution) and using Bayes rule, the *a posteriori* conditional probability density is given by:

$$\Pr_{t|r}(\tau \mid r) = \frac{\Pr_{r|t}(r \mid \tau) \Pr(\tau)}{\Pr(r)} == K \exp\left(\frac{1}{\sigma^2} \sum_t r(t) s(t-\tau) \right)$$

where K is a normalization factor that accounts for parts of the conditional density that do not depend on t. The Minimum Mean Square Estimator (MMSE) is the conditional expectation of a time delay given the observation:

$$\hat{t}_{MMSE} = E[\tau \mid r] = \sum \tau \Pr_{t|r}(\tau \mid r)$$

Since the evaluation of this sum involves computationally expensive and numerically unstable steps, the Maximum Likelihood Estimator (MLE) can be computed instead [20]. The MLE is obtained by maximizing log-likelihood function:

$$l(\tau) = \log \Pr_{r|t}(r \mid \tau) = K' + \frac{1}{\sigma^2} \sum_t r(t) s(t-\tau) = K' + \frac{C(\tau)}{\sigma^2}$$

This is equivalent to maximizing the cross-correlation (the output of the matched filter) between the received signal and the source signal [20]. Therefore, the Maximum Likelihood Estimator of time delay is

$$\hat{t}_{MLE} = argmax\ C(\tau)$$

Furthermore, the cross-correlation could be expressed in terms of the autocorrelation function of the source waveform and an additive filtered noise:

$$C(\tau) = \sum s(t - t_0) s(t-\tau) + \sum n(t) s(t-\tau) = R(\tau - t_0) + W(\tau)$$

Since the filtered noise is zero-mean, the signal autocorrelation $R(\tau)$ is the expected value of the cross-correlation $C(\tau)$ when averaged over all possible noise samples (the ensemble average likelihood function). As will be discussed later in this section, the multimodal shape of this ensemble-averaged log-likelihood function plays a critical role in appearance of outliers when the noise level rises above a certain

threshold. The amount of noise in the received signal is characterized by pre-filtered and post-filtered Signal-To-Noise Ratio (SNR) defined by

$$SNR_{pre} = \frac{P_s}{P_n} = \frac{\sum_t s(t)^2}{N\sigma^2},\ SNR_{post} = \frac{P_R}{P_W} = \frac{\sum_\tau R(\tau)^2}{\sigma^2 \sum_t s(t)^2}$$

Where P_s, P_n, P_R, P_W are the power of the source signal, the power of the noise, the power the autocorrelation of the source signal and filtered noise respectively. When the SNR level is high, the filtered noise $W(t)$ causes the MLE to be near the global maximum of the autocorrelation function at $\tau = t_0$. When the SNR falls below a certain threshold level SNR_{thresh}, the filtered noise may occasionally cause the global maximum of the received signal to be detected far away from the true value $\tau = t_0$.

For general autocorrelation function the "wrong" peak could emerge almost anywhere outside the vicinity of the global maximum of the autocorrelation function. But for a special class of signals and, more generally, for a special class of likelihood functions, the outliers will be clustered around certain locations that are characterized by the local maximum of the ensemble-average likelihood function. The narrowband signals that are frequently used in sonar and radar applications are good examples of such a class of signals since the autocorrelation function of narrowband signals has significant local maxima (side lobes).

Using discrete cosine representation, a DC-less narrowband source signal could be represented by

$$s(t) = \sum_{k=1}^{\infty} a_k \cos\left(2\pi f_k t + \phi_k\right)$$

The total (finite) signal energy, the central frequency, and mean square bandwidth of a signal in the above representation are given respectively by

$$E^2 = \sum_{k=1}^{\infty} a_k^2, F_c = \frac{1}{E^2}\sum_{k=1}^{\infty} a_k^2 f_k, B^2 = \frac{1}{E^2}\sum_{k=1}^{\infty} a_k^2 f_k^2$$

The centralized mean square bandwidth of a signal is given by

$$B_c^2 = \frac{1}{E^2}\sum_{k=1}^{+\infty} a_k^2 \left(f_k - F_c\right)^2 = B^2 - F_c^2$$

We assume that most of the signal energy is concentrated around its central frequency or, equivalently, the signal centralized bandwidth is much smaller than the signal central frequency, namely

$$\frac{B_c^2}{F_c^2} \ll 1 \text{ or } \left|\frac{f_k - F_c}{F_c}\right| = \left|\frac{f_k}{F_c} - 1\right| \ll 1 \text{ for all } k \text{ such that } a_k \neq 0$$

Under this assumption, the autocorrelation function of a signal has local maxima (side lobes) located at a distance $T_m = \pm m\frac{1}{F_c}$ from the peak at $t = 0$. For the rest of this section we consider only the two most significant side lobes corresponding to values of $m = \pm 1$ although in our simulations even more distant side

lobes are taken into account. The probability of outliers is the probability that a local maximum of the signal autocorrelation function becomes the global maximum of the cross-correlation function. Below we approximate the probability of this event (an outlier event) using a narrowband signal representation. Without loss of generality, we assume $t_o = 0$ as we are interested only in relative positions of the autocorrelation maxima. We also consider the signal and noise values only in the vicinity of the global maximum and near a single highest side lobe of the cross-correlation (two-point model). Then, using this model, the outlier probability is

$$\Pr(outlier) = \Pr\left(C(0) < C\left(\frac{1}{F_c}\right)\right) = \Pr\left(R(0) - R\left(\frac{1}{F_c}\right) < W\left(\frac{1}{F_c}\right) - W(0)\right)$$

Using Taylor expansion the left hand side of this inequality is approximated by

$$R(0) - R\left(\frac{1}{F_c}\right) = \sum_{k=0}^{\infty} a_k^2 \left((1 - \cos\left(2\pi \frac{f_k}{F_c}\right)\right) = \sum_{k=0}^{\infty} a_k^2 \left((1 - \cos\left(2\pi \frac{f_k - F_c}{F_c}\right)\right) \approx$$

$$\approx \frac{2\pi^2}{F_c^2} \sum_{k=0}^{\infty} a_k^2 \left(f_k - F_c\right)^2 = 2\pi^2 E^2 \frac{B_c^2}{F_c^2}$$

Denoting $V^2 = \frac{B_c^2}{F_c^2}$, the difference between values of autocorrelation function at its peaks could be expressed as $2\pi^2 E^2 V^2$. It is clear from the representation that the difference decreases when the central frequency is larger (as peaks are more closely spaced) or if the signal centralized bandwidth is lowered, (the signal envelope becomes less sharp). Asymptotically, if the signal bandwidth is made very small (tends to zero), the resulting signal approaches a sinusoid with zero differences between peaks height.

Now we turn our attention to the difference between filtered noise samples. Filtered noise samples are linear combinations of zero-mean normal variables and, therefore, are distributed as zero mean normal random variable with variance $\sigma_w^2 = E^2\sigma^2$. The covariance between filtered noise samples at distance $\frac{1}{F_c}$ is

$$Cov_W(0, \frac{1}{F_c}) = E\left[W(0) W\left(\frac{1}{F_c}\right)\right] = \sigma^2 R\left(\frac{1}{F_c}\right)$$

Again, using Taylor expansion the value of the autocorrelation function near the side lobe could be approximated by:

$$R\left(\frac{1}{F_c}\right) = \sum_{k=0}^{\infty} a_k^2 \left(\cos\left(2\pi \frac{f_k - F_c}{F_c}\right)\right) \approx E^2 \left(1 - 2\pi^2 V^2\right)$$

Therefore, the correlation coefficient between filtered noise samples at neighboring maxima is

$$\rho_W\left(0,\frac{1}{F_c}\right)\approx\frac{\sigma^2E^2\left(1-2\pi^2V^2\right)}{\sigma^2E^2}=1-2\pi^2V^2$$

The difference between two correlated normally distributed zero-mean random variables is also zero-mean normal with variance:

$$\sigma_{\triangle W}^2=2\sigma^2E^2-2\sigma^2E^2\left(1-2\pi^2V^2\right)=4\pi^2\sigma^2E^2V^2$$

Therefore, the probability of the outlier becomes

$$\Pr(\triangle R<\triangle W)=\Pr\left(\frac{\triangle R}{2\pi\sigma EV}<\frac{\triangle W}{2\pi\sigma EV}\right)=\Pr\left(Z>\pi V\frac{E}{\sigma}\right)=Q\left(\pi V\frac{E}{\sigma}\right)$$

Where Z is a random variable with standard normal distribution and Q is the tail probability of the standard normal distribution. Therefore, the probability of an outlier becomes large when either signal-to-noise ratio $\frac{E}{\sigma}$ or centralized bandwidth to central frequency ratio V is small.

For an outlier measurement, there will be a bias of $\pm\frac{1}{F_c}$ induced by left and right side lobes of the autocorrelation function. Since the probabilities of "left" and "right" outlier are equal, the overall bias due to outliers will disappear. However, since the impact of an outlier to the overall error is proportional to the bias squared, an outlier event will seriously affect MSE of the estimation. The resulting MSE could be decomposed into parts reflecting the impact from inlier and outlier events separately:

$$MSE=\Pr(outlier)\left(Variance_{outlier}^2+Bias_{outlier}^2\right)+(1-\Pr(outlier))Variance_{inlier}^2$$

Assuming that expected additional variance due to outlier

$$\Pr(outlier)\left(Variance_{outlier}^2-Variance_{inlier}^2\right)$$

is neglectable, the MSE is approximated by:

$$MSE\approx Variance_{inlier}^2+\Pr(outlier)\,Bias_{outlier}^2$$

Therefore, the cost of outlier for the overall MSE is proportional to

$$\frac{Q\left(Z>\pi V\frac{E}{\sigma}\right)}{F_c^2}$$

The analysis above is consistent with the well-known s-shaped curve describing dependency between the MSE of MLE and the Signal-to-Noise Ratio (SNR) [23]. For large values of SNR, the probability of outlier is small and, therefore, the MSE approaches the CRLB curve and the system is said to be in the coherent-state. As soon as SNR drops below a certain threshold value, the MSE increases abruptly due to the presence of outliers and, accordingly, the system is said to be in the semi-coherent state. For very low SNR values, the sidelobes produced by the strong presence of the central frequency in the signal spectrum do not play a significant role in the location of outliers and the system is said to be in non-coherent state.

For the rest of this analysis, we are mainly concerned with the semi-coherent area and the methods for improving the MSE by reducing the effect of outliers on it. We conclude this section by noting that the central frequency appears in both nominator and denominator of the expression above. Therefore, although the signals with higher central frequency have greater probability for outlier, the impact on overall MSE of each outlier event is smaller.

III. PHASE-SHIFTED UNMATCHED FILTERS

In this section, we introduce the Phase-Shifted Unmatched Filters and the associated biased ToF estimators. We extend the analysis from the previous section to show that although each of the biased ToF estimators has larger estimation error, the errors of these estimators are not completely correlated. This property, which we denote by Semi-correlated Estimators (SME) enables to combine (fuse) multiple biased ToF estimators into a single robust ToF estimator using machine learning algorithms which are described in the next section

A Phase-Shifted Unmatched Filter is generated from the source waveform $s(t)$ by shifting a phase of its entire spectrum content by a same value φ. The phase-shifting operation can be easily performed using the Hilbert Transform [7] that could be computed numerically. Continuing with previously used notation, the phase-shifted signal is represented as

$$s_{\varphi}(t) = \sum_{k=0}^{\infty} a_k \cos\left(2\pi f_k t + \phi_k + \varphi\right)$$

The cross-correlation between the received signal and the unmatched filter could be split into a sum of two terms as in the previous section:

$$C_{\varphi}(t) = R_{\varphi}(t) + W_{\varphi}(t)$$

The first term is the cross-correlation between the source signal and the unmatched filter. The second term represents an additive filtered noise. The cross-correlation between the source signal and unmatched filter is given by

$$R_{\varphi}(t) = \sum_{k=0}^{\infty} a_k^2 \cos\left(2\pi f_k t - \varphi\right)$$

The associated ToF estimator is obtained by locating the global maximum in the unmatched filter output, that is

$$\hat{t}_{\varphi} = \arg\max C_{\varphi}(t)$$

For small values of phase shift, the global maximum will be located near the peak of autocorrelation function. Therefore, the position of the global maximum of $R_{\varphi}(t)$ could be estimated by setting the time derivative to zero and using Taylor expansion around zero phase and time

$$\frac{\partial R_\varphi(t)}{\partial t} = -\sum_{k=0}^{\infty} a_k^2 2\pi f_k \sin\left(2\pi f_k t - \varphi\right) = 0$$

$$\sum_{k=0}^{\infty} a_k^2 f_k \sin\left(2\pi f_k t - \varphi\right) \approx \sum_{k=0}^{\infty} a_k^2 f_k \left(2\pi f_k t - \varphi\right) = E^2\left(2\pi B^2 t - F_c\varphi\right) = 0$$

and, therefore, the position of global maximum is

$$t_\varphi^M = \frac{\varphi}{2\pi}\frac{F_c}{F_c^2 + B_c^2} = \frac{\varphi}{2\pi}\cdot\frac{1}{F_c}\cdot\frac{1}{1+V^2}$$

The value of the cross-correlation function at its global maximum could be approximated using linear and quadratic terms of the Taylor expansion

$$R_\varphi(t_\varphi^M) = \sum_{k=0}^{\infty} a_k^2 \cos\left(\varphi\left(f_k\frac{F_c}{B^2} - 1\right)\right) \approx \sum_{k=0}^{\infty} a_k^2\left(1 - \frac{\varphi^2}{2}\left(f_k\frac{F_c}{B^2} - 1\right)^2\right)$$

$$\sum_{k=0}^{\infty} a_k^2\left(1 - \frac{\varphi^2}{2}\left(f_k\frac{F_c}{B^2} - 1\right)^2\right) = E^2 - \frac{\varphi^2}{2}E^2\left(\frac{F_c^2}{B^2} - 2\frac{F_c^2}{B^2} + 1\right) =$$

$$= E^2\left(1 - \frac{\varphi^2}{2}\cdot\frac{V^2}{1+V^2}\right) = R(0) - \frac{\varphi^2}{2}\cdot\frac{V^2}{1+V^2}\cdot E^2$$

The height of the peak of the cross-correlation between source signal and phase-shifted unmatched filter is lower compared to the peak of the source signal autocorrelation. Therefore, the unmatched filter does not maximize peak signal-to-noise ratio as the matched filter does. The difference $\varphi^2\frac{V^2}{1+V^2}E^2$ depends on a specific value of the phase shift. Since the phase shift appears in quadratic form, the difference is the same for positive and negative phase shift of the same magnitude.

The heights of two closest sidelobes of cross-correlation function which are located at distances $T_m = \pm\frac{1}{F_c}$ from its main peak could be estimated using Taylor expansion:

$$R_\varphi\left(t_\varphi^M \pm \frac{1}{F_c}\right) = \sum_{k=0}^{\infty} a_k^2 \cos\left(\varphi\left(f_k\frac{F_c}{B^2} - 1\right) \pm \frac{2\pi f_k}{F_c}\right) = \sum_{k=0}^{\infty} a_k^2 \cos\left(\varphi\left(f_k\frac{F_c}{B^2} - 1\right) \pm \frac{2\pi(f_k - F_c)}{F_c}\right)$$

$$R_\varphi\left(t_\phi^M \pm \frac{1}{F_c}\right) \approx \sum_{k=0}^{\infty} a_k^2\left(1 - \frac{1}{2}\left(\varphi\left(f_k\frac{F_c}{B^2} - 1\right) \pm \frac{2\pi(f_k - F_c)}{F_c}\right)^2\right)$$

$$R_\varphi\left(t_\phi^M \pm \frac{1}{F_c}\right) \approx R\left(\pm\frac{1}{F_c}\right) - \frac{V^2E^2}{1+V^2}\left(\frac{\varphi^2}{2} \pm 2\pi\varphi\right)$$

Therefore, the difference between the heights of sidelobes and the main peak is

$$\triangle R_\varphi\left(\pm\frac{1}{F_c}\right) = R_\varphi(t_\phi^M) - R_\varphi\left(t_\phi^M \pm \frac{1}{F_c}\right) = \triangle R\left(\pm\frac{1}{F_c}\right) \pm 2\pi\varphi E^2\frac{V^2}{1+V^2} = 2\pi^2E^2V^2\left(1 \pm \frac{\varphi}{\pi}\frac{1}{1+V^2}\right)$$

Thus, the difference between the heights of sidelobes depends on their location relative to the main peak. For positive phase shift, the difference between the main peak and the right sidelobe is greater than difference between the main peak and the left sidelobe (**figure 1**).

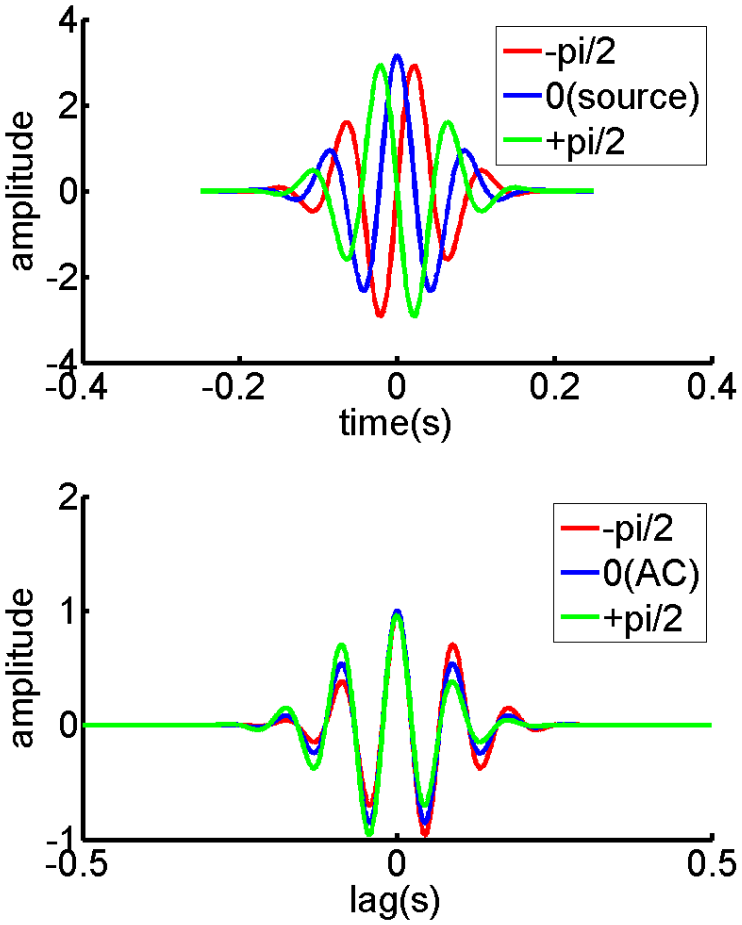

Figure 1: Two symmetrically phase shifted unmatched filters and the source signal (top); the time-aligned cross correlation between the source signal and unmatched filters (bottom)

Since the filtered noise has the same autocorrelation function as in case of the matched filter, the probability of a left outlier for $\hat{t}_\varphi$ is greater than the probability of the right outlier. For instance, the probability of the left outlier generated by the larger sidelobe is therefore

$$\Pr_\varphi(outlier_{left}) = \Pr\left(\triangle W\left(-\frac{1}{F_c}\right) > \triangle R_\varphi\left(-\frac{1}{F_c}\right)\right) = \Pr\left(Z > \frac{2\pi^2 E^2 V^2\left(1-\frac{\varphi}{\pi}\frac{1}{1+V^2}\right)}{2\pi\sigma E V}\right) =$$

$$= \Pr\left(Z > \pi V \frac{E}{\sigma}\left(1-\frac{\varphi}{\pi}\frac{1}{1+V^2}\right)\right) = Q\left(\pi V \frac{E}{\sigma}\left(1-\frac{\varphi}{\pi}\frac{1}{1+V^2}\right)\right)$$

It seems that the probability of at least one class of outliers is greater for a single phase-shifted estimator. However, the relation between probabilities of left and right outlier is reversed if we consider the estimator

generated by phase shift with the same magnitude but different sign. Therefore, the expected bias of two estimators with opposite phase shift is canceled out and their average constitutes an unbiased estimator for ToF.

$$\hat{t}_{\pm\varphi} = \frac{\hat{t}_{+\varphi} + \hat{t}_{-\varphi}}{2}$$

We have reported the improvement achieved by this simple average in [2]. The gain in performance obtained by such estimator depends on the SNR level and on a selected phase shift value. The employed value of phase shift defines the tradeoff in MSE between the inlier variance and the expected square of outlier bias. The optimal value of a phase shift could be calibrated to match the underlying SNR level and the source signal parameters [3]. Unfortunately, this method requires the knowledge of the operating SNR level which is not always available in practice. This requirement could be avoided by combining information from several pairs of symmetrically phase-shifted unmatched filters. Since each phase-shifted unmatched filter perturbates the noise signal in a slightly different way, each estimator derived from a different phase-shifted unmatched filter contains some additional information on the interrogated signal. For instance, for two different values of phase shift, the covariance between noise samples at global maximum is

$$Cov_w(W_\varphi(t_\varphi^M)W_\theta(t_\theta^M)) = \sigma^2 R_{\theta-\varphi}(t_\theta^M - t_\varphi^M) \approx \sigma^2 R_{\theta-\varphi}\left(\frac{\theta-\varphi}{2\pi}\cdot\frac{1}{F_c}\cdot\frac{1}{1+V^2}\right) \approx \sigma^2 E^2\left(1-\frac{(\theta-\varphi)^2}{2}\cdot\frac{V^2}{1+V^2}\right)$$

And, therefore, the correlation coefficient (assuming small value of phase shift)

$$\rho_{\varphi,\vartheta}^M = \left(1-\frac{(\theta-\varphi)^2}{2}\cdot\frac{V^2}{1+V^2}\right) < 1$$

Thus, for low SNR (presence of outliers) the phase shift of the source waveform generates a family of unmatched filters, from which a family of biased semi-correlated estimators can be constructed. This family of SMEs can be fused into a single estimator with improved estimation accuracy.

Since the correlation between outlier events is decreasing as the differences between phase shift values is increasing, it seems that, fusing a pair of estimators corresponding to a greater difference between corresponding phase shift values should make sense. However, as the underlying unmatched filters increasingly diverge from the shape of the matched filter, the error of the individual biased estimators increases. Therefore, since the probability of outliers increases with a decrease in the SNR, the estimators with larger phase shift differences should be assigned a greater weight in the SMEs fusion as the correlation between them becomes more important. This suggests that the fusion of the SMEs depends on the SNR.

In addition, since often an unbiased estimator is required, a fusion method should somehow reduce the combined bias of multiple input estimators to produce an unbiased output estimate.

In [4] we have introduced a method for fusion of multiple biased estimators using regression. The Decision Trees based regression [1] was trained using a mix of samples for a range of SNR values in the semi-coherent zone and, thus, the combined estimate did not require *a priori* knowledge of the operational SNR value. The method provided up to 10% improvement in the estimation accuracy compared to the conventional Maximum Likelihood Estimator.

In the next section, we introduce an alternative fusion scheme that relies on phase-shifted unmatched filters for a weak classification of ToF MLE outcomes into classes corresponding to peaks of the signal autocorrelation function. Using the labels produced by the classifier, the expected bias introduced by an outlier measurement is reduced from the MLE value, resulting in up to 30% increase in the estimation accuracy. The additional information generated by the classifier could be used also for devising an efficient method for fusion of multiple independent ToF measurements in the presence of outliers.

IV. MACHINE LEARNING FUSION OF BIASED ESTIMATORS

This section describes a novel machine learning method for improving the MSE of the Maximum-Likelihood ToF estimator in the semi-coherent region. The method employs a weak classifier that is trained to label the MLE value according to the side lobe of the autocorrelation function that induced the peak in the output of the matched filter. The classifier relies on a bank of phase-shifted unmatched filters to generate an input feature vector for training and classification. Based on the label produced by the classifier, an expected bias due to outliers can be computed using the estimated prior probabilities of different outlier types and the estimated confusion matrix of the classifier. The resulting estimator is obtained by subtracting the expected bias from the ToF MLE value. Below we describe the method in a general form to demonstrate its applicability to similar estimation problems.

Consider an estimator $\hat{x}$ of a parameter x contaminated by additive random noise $\eta : \hat{x} = x + \eta$. We model outliers by assuming that the random noise is sampled form m different distribution (similar to a mixture model). These noise probability distributions are characterized by the vector of means $\overline{\mu} = (\mu_1 ... \mu_m)$ and the vector $\overline{\sigma^2} = (\sigma_1^2 ... \sigma_m^2)$ of their variances. There is also a vector of prior probabilities $\overline{p} = (p_1 ... p_m)$ $\sum_{i=1}^{m} p_i = 1$ such that the probability of a noise sample to be selected from distribution i is p_i. We say that $N(\hat{x}) = i$ if noise at measurement $\hat{x}$ is generated using the probability distribution with mean μ_i and variance σ_i^2

The expected bias of the measurement is $b_0 = E\eta = \sum_{i=1}^{m} p_i \mu_i$. Therefore, we can compute the error of unbiased estimator $\hat{x}_u = \hat{x} - b_0$ as

$$e_0^2 = E(\hat{x}_u - x)^2 = E(\eta - b_0)^2 = \sum_{i=1}^{m} p_i \sigma_i^2 + \sum_{i=1}^{m} p_i \mu_i^2 - b_o^2$$

If the expected bias is zero as in the case of symmetric outliers induced by an autocorrelation function, the last term disappears and the total error becomes:

$$e_0^2 = \sum_{i=1}^{m} p_i \sigma_i^2 + \sum_{i=1}^{m} p_i \mu_i^2$$

Assume we can train a (weak) classifier that assigns labels $L(\hat{x}) = 1...m$ to an estimate $\hat{x}$ such that $\Pr(L(\hat{x}) = i \mid N(\hat{x}) = j) = g_{i,j}$ which is the probability of a measurement receiving a label i by the classifier given that the measurement's noise is generated by the j-th distribution. These conditional probabilities form the confusion matrix G of the classifier: $G = \{g_{i,j}\}_{i=1,j=1}^{m,m}$.Than we can express posteriori probabilities for a sample to be labeled as class i as:

$$q_i = \Pr(L(\hat{x}) = i) = \sum_{j=1}^{m} \Pr(L(\hat{x}) = i \mid N(\hat{x}) = j) \Pr(N(\hat{x}) = j) = \sum_{j=1}^{m} g_{i,j} p_j$$

or in a matrix form $\overline{q} = G \cdot \overline{p}$. Given this posteriori class probabilities, the likelihood matrix $C = \{c_{i,j}\}_{i=1,j=1}^{m,m}$ becomes:

$$c_{i,j} = \Pr(N(\hat{x}) = i \mid L(\hat{x}) = j) = \frac{\Pr(L(\hat{x}) = j \mid N(\hat{x}) = i) p_i}{q_j} = \frac{g_{j,i} p_i}{q_j}$$

Based on this likelihood matrix, we can also compute the conditional bias of a measurement given the classifier has labeled it with class i .

$$b_i = E_{L(\hat{x})=i}(\hat{x} - x) = E_{L(\hat{x})=i}\eta = \sum_{j=1}^{m} \mu_j c_{j,i}$$

Therefore, we can reduce appropriate bias b(x) from the estimate using additional information obtained from the classifier using:

$$b(\hat{x}) = b_i \text{ if } L(\hat{x}) = i$$

The expected value of this function is equal to the bias of the original estimator as

$$Eb(\hat{x}) = \sum_{i=1}^{m} q_i b_i = \sum_{i=1}^{m} q_i \sum_{j=1}^{m} \mu_j c_{j,i} = \sum_{j=1}^{m} \mu_j p_j = b_o$$

This allows constructing the modified unbiased estimator $\hat{x}_m = \hat{x} - b(\hat{x})$. The MSE of the new modified estimator is

$$e_m^2 = E(\hat{x}_m - x)^2 = E(\eta - b(\hat{x}))^2 = E\eta^2 - \sum_{i=1}^{m} q_i b_i^2 = e_o^2 + b_o^2 - \sum_{i=1}^{m} q_i b_i^2 = e_o^2 + \left(\sum_{i=1}^{m} q_i b_i \right)^2 - \sum_{i=1}^{m} q_i b_i^2$$

The improvement of the MSE depends on the differences between the last two terms which is always non-positive due to Jensen's inequality [14]

$$\left(\sum_{i=1}^{m} q_i b_i \right)^2 \le \sum_{i=1}^{m} q_i b_i^2$$

For the case of the symmetric outliers $b_o = \sum_{i=1}^{m} q_i b_i = \sum_{i=1}^{m} p_i \mu_i = 0$ and for a perfect classification achieving $q_i = p_i$, the bias square part of the MSE is completely reduced, resulting in MSE containing only the variance part

$$e_m^2 = \sum_{i=0}^{m} p_i \sigma_i^2$$

To demonstrate the accuracy gain obtained by this approach, we consider a simple example involving two symmetric outliers (called left and right outliers) that occur with prior probability $p_O \le \frac{1}{2}$ each. Therefore, the vector of the prior probabilities becomes:

$$\overline{p} = (p_L, p_{M,} p_R) = (p_{O,} p_I = 1 - 2p_O, p_O).$$

The outliers are symmetric, thus, the vector of means is $\overline{\mu} = (\mu_L, 0, \mu_R) = (-\mu_O, 0, \mu_O)$. The variances of outliers are identical and are different from the variance of the inlier $\overline{\sigma}^2 = (\sigma_O^2, \sigma_I^2, \sigma_O^2)$. Assume we have managed to train a classifier which produces a constant misclassification rate on all classes. That is

$$g_{i,j} = \Pr(L(\hat{x}) = i \mid L(\hat{x}) = j) = \varepsilon \le \frac{1}{2}, \forall i \ne j \text{ and}$$

$$g_{i,i} = \Pr(L(\hat{x}) = i \mid L(\hat{x}) = i) = 1 - 2\varepsilon = \delta .$$

Under these settings, the bias b_0 of the estimator $\hat{x}$ is zero and the error becomes

$$e_0^2 = p_I \sigma_I^2 + 2p_O \sigma_O^2 + 2p_O \mu_O^2.$$

The vector of *posterior* class probabilities can be shown to be

$$\overline{q} = (q_L, q_M, q_R) = (q_O, q_I, q_O) = (p_O + \varepsilon(1 - 3p_O), p_I \delta + 2p_O \varepsilon, p_O + \varepsilon(1 - 3p_O))$$

And finally, the conditional expected biases given the class labels are

$$b_M = 0$$

$$b_L = -b_R = \frac{\mu_O p_O}{q_L}(\varepsilon - \delta)$$

Then the improvement in the MSE is equal to

$$\triangle e^2 = e_o^2 - e_m^2 = \sum_{i=1}^{m} q_i b_i^2 = \frac{2\mu^2 p_O^2}{q_O}(\varepsilon - \delta)^2 = 2\mu^2 \frac{p_O^2 (1 - 3\varepsilon)^2}{\varepsilon + p_O (1 - 3\varepsilon)}$$

Again, for a perfect classifier with $\varepsilon = 0$, the difference achieves its maximum at $\triangle e^2 = 2\mu^2 p_O$ and, therefore, it completely cancels out the bias square part in the estimator error. For random guess classifier with $\varepsilon = \frac{1}{3}$, this difference becomes zero and no gain in the accuracy can be achieved.

The above algorithm requires parametric modeling of outlier classes to obtain estimates on μ,σ and the prior probabilities p. These parameters could be obtained from labeled training data by various methods (e.g. EM algorithm). The same training data could be used to train multiclass classifier to assign labels according to outlier classes. In addition, the classifier performance needs to be estimated and summoned in a confusion matrix which might be done using the "evaluation" data set.

In the next section, we describe the training and estimation process along with simulations that are used for evaluating the method. The simulations show that the method achieves a significant improvement (up to 30%) in the accuracy of the semi-coherent ToF estimation.

V. THE ESTIMATION PROCESS

The simulation process (**Figure 2**) consists of two phases. During the preprocessing phase (blocks P1 through P4 in the diagram), a large data set of randomly generated samples is created and then different portions of this data set are used for training the classifier and for the evaluation of the classifier performance. During the estimation phase (blocks E1 through E4 in the diagram), the ToF estimation for a new simulated sample is computed using the classifier's predictions.

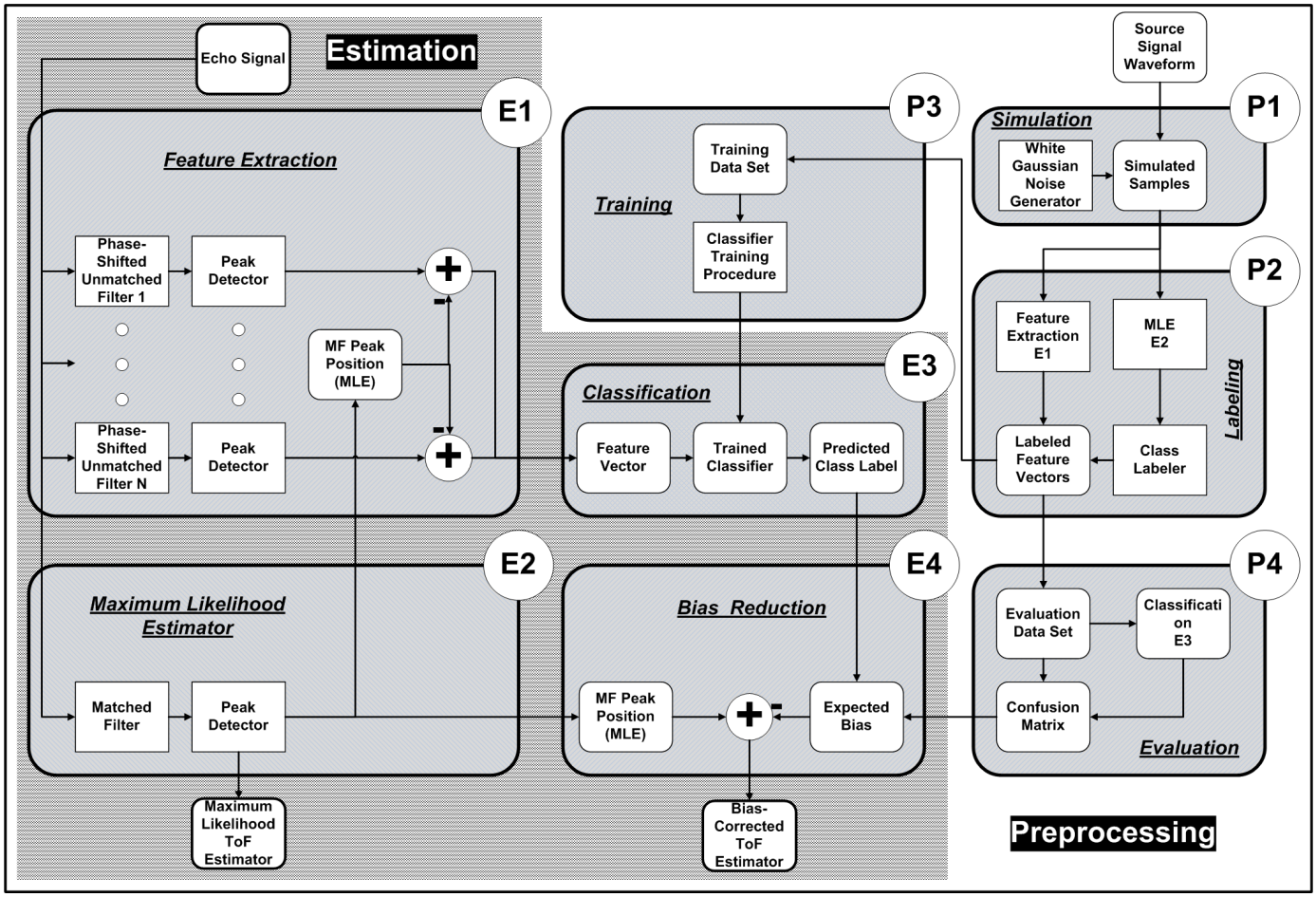

Figure 2: The simulation process consists of the preprocessing phase (P1-P4) and the estimation phase (E1-E4). During the preprocessing phase the training and the evaluation data sets are

generated (P1) and, then, are labeled by appropriate inlier/outlier class labels (P2). The training data set is used for training a weak classifier (P3) and the evaluation data set is used for estimating the classifier's confusion matrix (P4). During the estimation phase, the feature vector is extracted from the sample (E1) and fed into the classifier (E3). The predicted class label is used for computing the expected bias and subtracting it (E4) from the previously computed Maximum Likelihood Estimator (E2).

We start with the description of the preprocessing phase. About 10000 data samples each representing a received signal are simulated by adding random White Gaussian Noise (WGN) to the selected narrowband signal (P1 at the diagram). The noise is produced using ten post-filtered SNR values corresponding to the semi-coherent range (3-8dB). The samples of all SNR values are mixed together for all further processing. A feature vector for each sample is extracted using a bank of n phase shifted filters (similar to the feature extraction step performed during the estimation phase). The filters in the bank are generated using the original source signal waveform. For an odd n, the values of n phase values are selected from the interval $[a:a]$ according to

$$\varphi_k = -a + 2a\frac{k-1}{n-1} \quad \text{where } k = 1,...,n$$

Thus, $\varphi_k = -\varphi_{n-k}$ and $\varphi_{\frac{n+1}{2}} = 0$, meaning that the middle element in the feature vector corresponds to the Maximum Likelihood Estimator. An n-valued feature vector $\overline{t} = (t_{\varphi_i},...,t_{\varphi_n})$ is formed by computing the locations of the global maximum in n cross-correlations:

$$t_{\varphi_i} = \arg\max C_{\varphi_i}(t),$$

where $C_{\varphi_i}(t)$ is the cross-correlation of *i*th filter with a simulated sample. The value returned by the Maximum Likelihood Estimator (which is obtained using the zero phase-shifted filter, aka the matched filter) is subtracted from all elements of $\overline{t}$. Therefore, a corrected feature vector is computed as

$$\overline{t}^{\,c} = (t^c_{\varphi_1},...,t^c_{\varphi_n}) = (t_{\varphi_1} - t_{\varphi_{\frac{n+1}{2}}},...,t_{\varphi_n} - t_{\varphi_{\frac{n+1}{2}}}) = (t^c_{\varphi_1},...,t^c_{\varphi_{\frac{n-1}{2}}},0,t^c_{\varphi_{\frac{n+3}{2}}},...,t^c_{\varphi_n})$$

The subtraction of the MLE value is necessary as we are interested in identifying an outlier by its locations relatively to the location of the global maximum of the autocorrelation function. Consequently, n-valued feature vectors always will have zero as their middle element and therefore they actually represent total of n-1 non-trivial features.

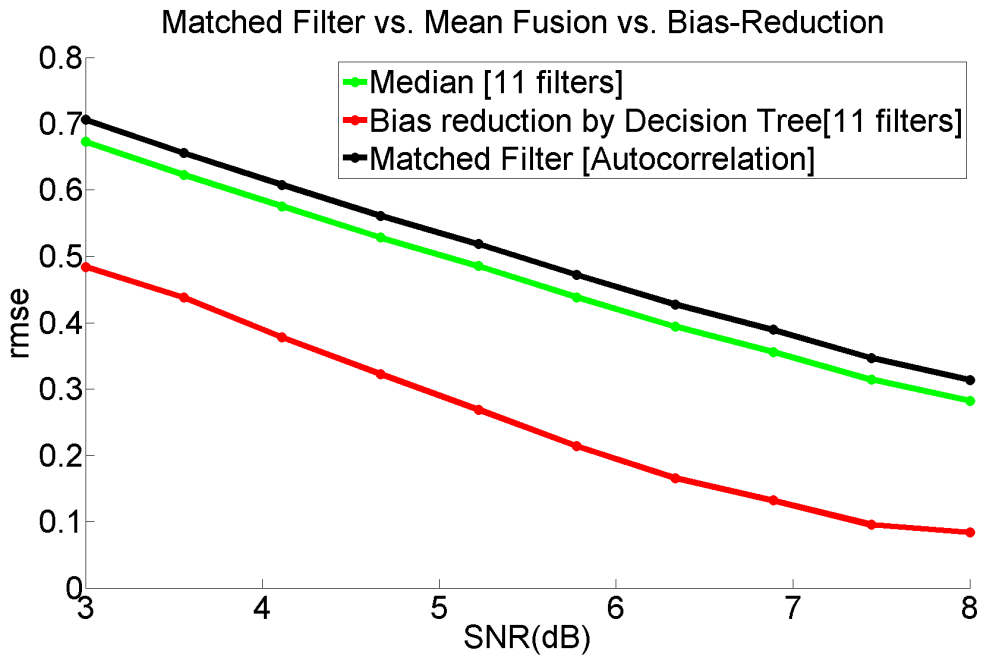

Figure 3: The RMSE of the ToF MLE (black), the bias-corrected estimator (red) and of the estimator obtained by the fusion of individual biased estimators using median statistics (green). The median-based estimator gives slight improvement over the MLE by discarding the extreme values. The bias-corrected ToF estimators improve the classical MLE by about 30%

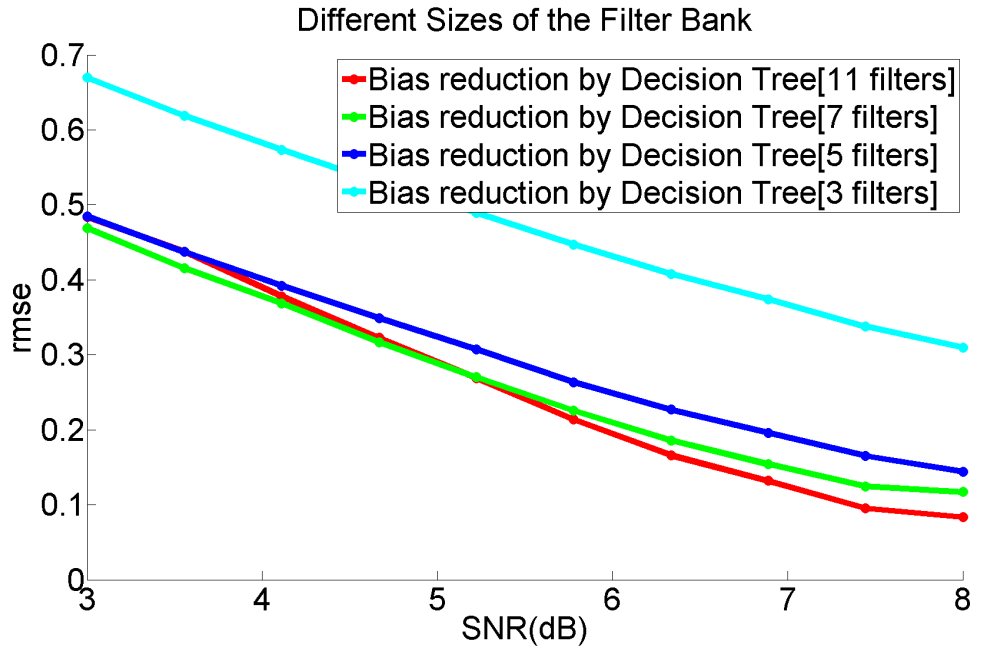

Figure 4: The RMSE obtained using filter banks of different sizes. The n=11 (red) gives better performance at higher SNR as more delicate interrogation of the signal required. On the other hand, it results in larger generalization error compared to more course estimator for n=7 (green)

The MLE values are also used for computing the samples' "true" labels according to the distance between the MLE and the true value of time-delay. One out of m possible class labels is assigned to a

sample according to its distance to one out of m selected peaks of autocorrelation function. Then, about a half of the labeled data is used to train the decision trees classifier (block P3 in the diagram) which accept n-valued feature vector and produces one out of m possible class labels. The second half of the data is used for evaluating the performance of the classifier, that is for computing the confusion matrix and the posteriori class probabilities (block P4 in the diagram).

The results presented in the next section are obtained using values of $a = \frac{\pi}{2}$ and $n = 11, 7, 3$. The number of classes - m is set to five with four outlier classes corresponding to four largest sidelobes of the autocorrelation function (two at each side of the main lobe) and the inlier class corresponding to the main lobe of the autocorrelation function.

After the preprocessing phase, the ToF estimation for a new sample proceeds in straightforward manner. The feature vector for a sample is extracted by applying n phase-shifted filters and computing corresponding biased estimates (block E1 in the diagram). After the feature vector is corrected by subtracting the MLE value, the vector is fed into the classifier which produces one out of m possible class labels (block E3 in the diagram). The expected bias is then computed as described in the previous section using the classifier's confusion matrix and posteriori class probabilities. The computed expected bias is then subtracted from the MLE resulting in a bias-corrected ToF estimate.

For comparison purposes, the simpler fusion of individual biased estimates is computed using a median statistics as

$$t_{median} = median\{t_{\varphi_1}, ..., t_{\varphi_n}\}$$

The resulting Root Mean Square Error (RMSE) normalized by the signal central frequency is shown in **figure 3.** The filter bank of n=11 filters is used in this experiment. From the graph, it is clear that the resulting error is significantly reduced (up to 30%) by correcting the MLE estimator using the expected bias (provided the estimation is carried-out in the semi-coherent area). The median-based fusion does not provide similar improvement although it manages to reduce effect of outliers by discarding some extreme values. **Figure 4** shows RMSE curves obtained using different sizes of filter banks. Generally, the more filters are employed, the more delicate interrogation of the signal is possible. Since samples with all SNR levels are mixed together for training, a finer analysis of the signal results in a larger generalization error. This is illustrated in **Figure 4** by the n=11 line having better performance over the n=7 line in the region of higher SNR levels. The relation is reversed as more "coarse" n=7 case gives smaller generalization error for lower SNR values.

VI. ESTIMATION FROM MULTIPLE PULSES

In this section, we introduce a method for estimating near-the-threshold ToF from multiple measurements. The method employs weighted averaging of individual estimates with carefully chosen weights which reflect the degree of uncertainty associated with each estimates.

The effect of the outliers on the overall accuracy of a ToF measurement is twofold. First, the MSE error increases due to the squared bias component introduced by an outlier. This effect was treated by reducing the expected bias from the MLE measurement as described above and in more details in [4]. The second impact of an outlier to the estimation accuracy is due to the increased uncertainty associated with outlier measurements. Although the outlier measurements are clustered around the local maxima of the autocorrelation function, the spread of the outlier estimates around the local maxima is significantly greater compared to the spread of inlier events around the central peak of the autocorrelation. This increase in the uncertainty (or intra-class variance) for outlier measurements could be explained by considering the Signal-to-Noise ratio near a side lobe of the autocorrelation function. The expected power of the filtered noise sample is the same near the side lobes as near the main peak. However, the signal power near the sidelobes are significantly lower leading to wider spread of the measurements near a lower peak. Below we introduce the Optimal Weighted Averaging method for combining (fusing) multiple independent estimates in the presence of outliers. The method could be applied to other estimation problems besides ToF estimation and, thus, presented in general form to stress its wider applicability which is described further in the conclusion section.

Consider an estimator $\hat{x}$ of a parameter x. The effect of noise on the estimator is modeled by additive random variable η added to the true value of the parameter

$$\hat{x} = x + \eta .$$

Multiple outlier classes corresponding to different side lobes of autocorrelation function are modeled by assuming that the random noise is sampled from m different distribution (similar to a mixture model). These noise probability distributions are characterized by the vector $\bar{\mu} = (\mu_1 ... \mu_m)$ of their means and the vector $\overline{\sigma^2} = (\sigma_1^2 ... \sigma_m^2)$ of their variances. There is also a prior probabilities vector $\bar{p} = (p_1 ... p_m) \sum_{i=1}^{m} p_i = 1$ such that the probability of a noise sample to be selected from distribution i is p_i. We say that $T(\hat{x}) = i$ if the noise at measurement $\hat{x}$ is generated using the probability distribution with a mean μ_i and the σ_i^2. Without loss of generality, we assume that the *a priori* bias of the measurement $b^0 = E\eta = \sum_{i=1}^{m} p_i \mu_i$ is zero.

Therefore, we can compute the error of the estimator

$$e_0^2 = E(\hat{x}_u - x)^2 = E\eta^2 = \sum_{i=1}^{m} p_i \sigma_i^2 + \sum_{i=1}^{m} p_i \mu_i^2$$

Assuming we can train a (weak) classifier that assigns labels $L(\hat{x}) = 1 ... m$ to an estimate $\hat{x}$ such that $\Pr(L(\hat{x}) = i \mid T(\hat{x}) = j) = g_{i,j}$. $g_{i,j}$ is the probability for a measurement to receive a label i given that the measurement's noise is generated by j-th distribution. These conditional probabilities form the confusion matrix G of the classifier: $G = \{g_{i,j}\}_{i=1,j=1}^{m,m}$. Than we can express posteriori probabilities for a sample to be labeled as class i as:

$$q_i = \Pr(L(\hat{x}) = i) = \sum_{j=1}^{m} \Pr(L(\hat{x}) = i \mid T(\hat{x}) = j) \Pr(T(\hat{x}) = j) = \sum_{j=1}^{m} g_{i,j} p_j$$

Given these posteriori class probabilities, the likelihood matrix $C = \{c_{i,j}\}_{i=1,j=1}^{m,m}$ becomes:

$$c_{i,j} = \Pr(T(\hat{x}) = i \mid L(\hat{x}) = j) = \frac{\Pr(L(\hat{x}) = j \mid T(\hat{x}) = i)\, p_i}{q_j} = \frac{g_{j,i} p_i}{q_j}$$

For N independent unbiased measurements (estimates) $\hat{x}_1,..,\hat{x}_N$ the unbiased estimator obtained by fusion of the individual estimators is

$$\hat{x}_A = \frac{1}{N}\sum_{j=1}^{N} \hat{x}_j$$

The Mean Squared Error of this estimator is

$$e_A^2 = E(\hat{x}_A - x)^2 = \frac{1}{N} e_o^2$$

However, the fusion of the individual measurements by a simple mean (averaging) statistic does not take into account additional information gained from the labels produced by the classifier. Obviously, the measurements which are suspected to be outliers should have less impact on the overall results compared to the inlier measurements. Therefore, we consider a weighted average as a method for fusion of individual bias-corrected measurements

$$\hat{x}_w = \sum_{j=1}^{N} w_j \left(\hat{x}_j - b(\hat{x}_j)\right)$$

$$\sum_{j=1}^{N} w_j = 1 \text{ , } \mathrm{w}_j \geq 0 \text{ j=1..N}$$

Where $b(\hat{x}_j)$ is the *a posteriori* correction for the expected bias given the classifier output (as described in the previous section) and in more detail in [4]. For the sake of simplicity, we assume that the *a posteriori* bias is eliminated using the method described in the previous section. Thus

$$b_i = E_{L(\hat{x})=i} b(\hat{x}_j) = \mu_i$$

The optimal fusion weights are computed in a manner that ensures smaller weights assigned for measurements identified by classifier as outliers. Given the class label assigned by the classifier to a measurement, we can compute *a posteriori* the conditional expected error of the measurement that received label i as:

$$e_i^2 = E_{L(\hat{x})=i}(\hat{x} - b_i - x)^2 = E_{L(\hat{x})=i}\left(\eta - b_i\right)^2 = E_{L(\hat{x})=i}\left(\eta - \mu_i\right)^2 =$$

$$= Var_{L(\hat{x})=i}(\eta) = \sum_{j=1}^{m} \sigma_j^2 \Pr(T(\hat{x}) = j \mid L(\hat{x}) = i) = \sum_{j=1}^{m} \sigma_j^2 c_{j,i}$$

To compute the optimal weights given labels assigned by a classifier to each measurement, we formulate an optimization problem as follows; Given N measurements, we are looking for the optimal vector of weights $\overline{w} = \{w_i\}_{i=1}^{N}$ such that $\sum_{i=1}^{N} w_i = 1$ and $w_i \geq 0$ for all i. The optimal weights should minimize the

expected square error of weighted average given labels $L=(l_1...l_N)$ assigned by classifier to each measurement. The mean squared error of weighted average of bias-corrected estimators is:

$$e_W^2 = E\left(\sum_{i=1}^{N}\left[w_i(\hat{x}_i - b_{l_i})\right] - x\right)^2 == E\left(\sum_{i=1}^{N} w_i\eta_i\right)^2 = \sum_{i=1}^{N} w_i^2 E_{L(\hat{x}_i)=l_i}\eta_i^{\,2} = \sum_{i=1}^{N} w_i^2 e_{l_i}^2$$

where $e_{l_i}^2$ is the conditional error provided measurement i received label l_i.

Using Lagrange multipliers, we formulate the target function:

$$f(\overline{w},\lambda) = \sum_{i=1}^{N} w_i^2 e_{l_i}^2 + \lambda(\sum_{i=1}^{N} w_i - 1)$$

Differentiating the target function above, we obtain the system of N+1 linear equations:

$$\begin{cases} 2w_i e_{l_i}^2 + \lambda = 0 \text{ for } i = 1..N \\ \sum_{i=1}^{N} w_i = 1 \end{cases}$$

Solving the system gives following expressions for the optimal weights:

$$w_i = -\frac{\lambda}{2e_{l_i}^2} = \frac{1}{\sum_{i=1}^{N}\frac{1}{e_{l_i}^2}} \cdot \frac{1}{e_{l_i}^2} = D\frac{1}{e_{l_i}^2}$$

where $D = \left(\sum_{i=1}^{N}\frac{1}{e_{l_i}^2}\right)^{-1}$ is normalization constant. Then, the overall estimation error for the weighted fusion

$$e_w^2 = D^2\sum_{i=1}^{N}\frac{e_{l_i}^2}{e_{l_i}^4} = D^2\sum_{i=1}^{N}\frac{1}{e_{l_i}^2} = D = \left(\sum_{i=1}^{N}\frac{1}{e_{l_i}^2}\right)^{-1}$$

Since the expected number of measurements that are labeled by classifier with label i is $q_i N$, the expected error for fusion of N estimates becomes

$$\overline{e}_w^2 = \left(\sum_{i=1}^{m}\frac{Nq_i}{e_i^2}\right)^{-1} = \frac{1}{N}\left(\sum_{i=1}^{m}\frac{q_i}{e_i^2}\right)^{-1}$$

While for non-weighted average of bias-corrected estimates, the total expected error is

$$e_A^2 = \frac{1}{N}e_o^2 = \frac{1}{N}\sum_{k=1}^{m} p_i\sigma_i^2 = \frac{1}{N}\sum_{i=1}^{m} q_i e_i^2$$

Due to Jensen's inequality $\left(\sum_{i=1}^{m} q_i e_i^2\right)^{-1} \le \sum_{i=1}^{m}\frac{q_i}{e_i^2}$ and, thus, $e_A^2 \ge e_W^2$. Therefore, the expected error of the weighted average is smaller than the error of the conventional fusion by the averaging.

The resulting optimal weights are inversely proportional to the conditional expected error given the label assigned to a measurement. Thus, the values that correspond to the outlier classes that generally have larger expected error will receive smaller weights. The simulation results in section IV shows a significant increase in the accuracy even compared to the fusion using the median. Better performance could be attributed to the more efficient utilization of available information. The mode and median fusion statistics essentially discard information present in the outlier measurements while simple averaging is not robust enough to absorb the effect of outliers. The fusion of bias-corrected estimators by the Optimal Weighted Averaging (OWA) combines the best of two worlds as it takes into account information contained in all measurements while reducing the error introduced by outliers' bias and uncertainty.

As an illustrative example, we consider a simple case of a single inlier and a single outlier class. Assuming a zero bias for each case, the conditional square errors are e_I^2 and e_O^2 for inlier and outlier classes respectively. Naturally, we expect the error produced by outlier to be significantly larger than the inlier error $e_I^2 \ll e_O^2$. Denoting by q_I the a posteriori probability that a measurement will be labeled as inlier by a classifier, the expected error for fusion by averaging becomes

$$e_A = \frac{1}{N}\left(q_I e_I^2 + (1-q_I)e_O^2\right)$$

and for weighted average fusion

$$e_W = \frac{1}{N}\left(\frac{q_I}{e_I^2} + \frac{1-q_I}{e_O^2}\right)^{-1}$$

Therefore, the gain achieved by employing weighted average instead of simple averaging is

$$G = \frac{e_A - e_W}{e_A}$$

After some simplification, the gain could be expressed as

$$G = \frac{q_I e_I^2 + (1-q_I)e_O^2 - \dfrac{e_I^2 e_O^2}{q_I e_O^2 + (1-q_I)e_I^2}}{q_I e_I^2 + (1-q_I)e_O^2} = \frac{q_I(1-q_I)(e_I^2 - e_O^2)^2}{\left(e_O^2 + q_I(e_I^2 - e_O^2)\right)\left(e_I^2 - q_I(e_I^2 - e_O^2)\right)}$$

Using following notation

$$\lambda = \frac{e_O^2}{e_I^2},\ \beta = q_I(1-q_I)$$

The expected gain can be simplified even further to obtain

$$G = \frac{q_I(1-q_I)(1-\lambda)^2}{(\lambda + q_I(1-\lambda))(1-q_I(1-\lambda))} = \frac{q_I(1-q_I)}{q_I(1-q_I) + \frac{\lambda}{(1-\lambda)^2}} = \frac{\beta}{\beta + \frac{\lambda}{(1-\lambda)^2}}$$

For a perfect classification, the posteriori class probabilities and class conditional error are equal to a priori class probabilities and class spread respectively. Therefore, for a perfect classification

$$G = \frac{p_I(1-p_I)}{p_I(1-p_I) + \frac{\sigma_o^2/\sigma_i^2}{(1-\sigma_o^2/\sigma_i^2)^2}}$$

which gives the best possible improvement for a two-class case. The improvement depends on the probability of outlier and the relation between spreads of outlier and inlier class. For instance, if the probability of inlier is $p_I = 2/3$ and the ratio between variances $\sigma_o^2/\sigma_i^2 = 2$ then we can ideally achieve 50% improvement in measurement accuracy. Of course, the actual gain would be smaller due to unavoidable misclassifications but could be even larger for other cases provided the outlier probabilities, variance ratios, or number of outlier classes is increased.

During the simulation process, a total number of 10000 samples are simulated by adding White Gaussian Noise (WGN) to the selected narrowband signal. The noise is generated using ten post-filtered SNR values in the range of 0-10dB which includes the semi-coherent range (3-8dB). A feature vector for each sample is extracted using a bank of 11 phase shifted unmatched filters. For each SNR value, an equal number of samples are mixed into the training, evaluation and test sets. During the data preparation step, each measurement is labeled with a class label (i.e. a peak in autocorrelation function selected by the MLE). These "true" labels are used to compute a priori statistics for each class.

The decision trees classifier is trained to predict an outlier class for a new measurement based on the vector of responses the biased estimators obtained using phase-shifted unmatched filters. Using the evaluation set, the confusion matrix for the classifier is estimated as described in the previous section. Using predicted labels, the classifier's confusion matrix and prior statistics, the expected conditional bias for each measurement is computed and then is subtracted from the Maximum Likelihood Estimator.

The weights of individual measurements for the fusion by Optimal Weighted Averaging are computed based on the labels produced by the classifier, the classifier confusion matrix, and prior statistics. The fusion of groups of n=3,5,10 measurements from the test set is performed using simple averaging (mean), robust statistic (median) and the Optimal Weighted Averaging method. The resulting Root Mean Square Error (RMSE) is presented in **Figure 5.** It could be seen that the fusion by median statistic produces a smaller error as compared to the conventional averaging. However, median-based fusion completely discards the information which is present in these measurements. The OWA fusion method significantly improves over the median-based fusion because it manages to extract much more information from the available data while correcting for expected bias and weighting according to the expected degree of uncertainty.

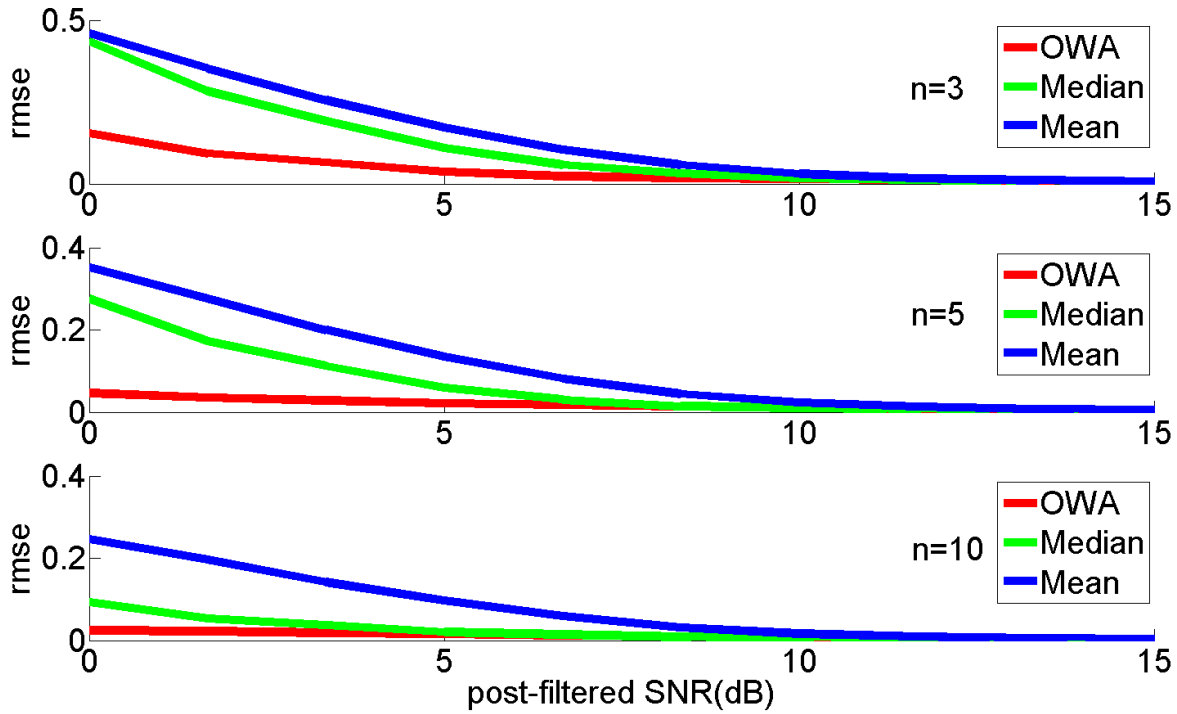

Figure 5: Root Mean Square Error of the ToF estimation as function of post-filtered SNR for fusing different number of measurements(n=3,5,10 measurements for top, middle and bottom figures respectively). The fusion by robust statistics (Median) discards outliers and, thus, does not provide significant benefits for fusion of small number of measurements (top). The optimal weighted averaging (OWA) fusion method weights individual measurements according to their degree of uncertainty and, thus, provides better results.

VII. THE ADAPTIVE SCHEME FOR CONTROLLING THE NUMBER OF MEASUREMENTS

In this section we describe a method for adaptively controlling number of measurements (pulses) required to achieve desired accuracy. The method employs the classifier described in the previous section combined with the early stopping rule which is introduced below.

For the sake of simplicity, in this section we assume only a single inlier and a single outlier class. However, the analysis below could be extended to multiple outlier classes. It is also assumed that $e_I^2 \ll e_O^2$. That is the conditional expected square error given the inlier label is significantly less than the error conditioned on an outlier label. We also assume that a posteriori probability of detecting an outlier is significantly less then posteriori probability for inlier detection $q_O = 1 - q_I \ll q_O$.

Then, the error of the weighted average estimator could be expressed as

$$e_{WA}^2 \approx \left(\frac{N_I}{e_I^2} + \frac{N_0}{e_O^2} \right)^{-1} = \frac{e_I^2}{N} \left(R + \rho(1-R) \right)^{-1}$$

where N_I, N_O is the number of inliers and outliers reported by the classifier, such that $N_I + N_O = N, R = \frac{N_I}{N}$ and $\rho = \frac{e_I^2}{e_O^2} \leq 1$ is the ratio of posteriori conditional errors. Without loss of generality we assume e_I^2=1 to eliminate the constant factor in the further analysis. Thus, the error for weighted average fusion becomes

$$e_{WA}^2 \sim \frac{1}{N_I + \rho N_O} = \frac{1}{N} \cdot \frac{1}{R + \rho(1-R)}$$

A conventional procedure for measurement consists of fixing the value of N (number of pulses), performing the measurements and, then, applying a fusion scheme using simple or weighted averaging or any other method for combining individual measurements. In the expression above, the number of pulses N and the ratio of errors ρ are completely predetermined by the measurement settings and the quality of the classifier respectively. In the conventional method, the only random factor which depends on a specific outcome (a series of N independent Bernoulli trials) is R. Without using a classifier, the actual inlier to outlier ratio is unobserved during the fusion (e.g. averaging), thus, making it impossible to treat each measurement according to its degree of uncertainty. Using a classifier, we can obtain some information regarding the certainty of an individual measurement in form of a class label. Although not perfect, this information allows us to process outcomes differently, resulting in a decrease in the expected error.

The OWA fusion method uses the labels computed by classifier at the end of the measurements process. However, the classifier could be applied after each measurement is made, providing additional information that could be used to for altering the measurement process itself. This can be achieved by stopping the measurements before reaching the target number of measurements N provided enough inlier measurements have been obtained. This early stopping rule can be illustrated by following simple example. Consider a measurement process where maximum of two pulses are used. However, if the first measurement is labeled by classifier as an inlier, the second measurement is not made. The expected number of measurements in this adaptive process is

$$EN_{adp} = q_I + 2(1-q_I) = 2 - q_I$$

The expected error is

$$Ee_{adp}^2 = q_I \frac{1}{1} + (1-q_I)\left(\frac{q_I}{1+\rho} + \frac{1-q_I}{2\rho}\right)$$

Since the adaptive method results in fractional number of expected pulses, we need to define a non-adaptive measurement process that uses the same number of expected pulses. In this way, we will be able to compare the error of adaptive and non-adaptive processes of the same average energy utilization.

We define the non-adaptive process by simply allowing flipping an unbalanced coin at the beginning of the process. With the probability q_I the measurement process proceeds by taking a single measurement and with the probability $1-q_I$ the measurement process uses 2 pulses (and twice as much energy). The difference between the adaptive and the non-adaptive schemes is very essential. In the non-adaptive scheme

the decision on number of pulses are made using prior to any measurements. In the adaptive scheme, a label produced by classifier on the first measurement determines if there is a need for the additional measurements. Then, the expected number of pulses used by the non-adaptive method is

$$EN_{1-2,wa} = q_I EN_{1,wa} + (1-q_I) EN_{2,wa} = 2 - q_I = EN_{adp}$$

The expected error of non-adaptive process is

$$Ee^2_{1-2,wa} = q_I Ee^2_{1,wa} + (1-q_I) Ee^2_{2,wa}$$

We are going to show that this error is larger than the error obtaining using adaptive scheme that is

$$\Delta = Ee^2_{1-2,wa} - Ee^2_{adp} > 0$$

Expanding, we obtain

$$\Delta = q_I(1-q_I)\left(\frac{1}{\rho}-1\right) + (1-q_I)\left(\frac{q_I^2}{2} + \frac{2q_I(1-q_I)-q_I}{1+\rho} + \frac{(1-q_I)^2-(1-q_I)}{2\rho}\right) =$$

$$= \frac{(1-q_I)\ q_I}{2}\left(\frac{1+\rho-2\rho^2}{\rho(1+\rho)} + q_I\frac{(1-\rho)^2}{\rho(1+\rho)}\right)$$

Therefore, we need to show that

$$\left(\frac{1+\rho-2\rho^2}{\rho(1+\rho)} + q_I\frac{(1-\rho)^2}{\rho(1+\rho)}\right) > 0$$

or

$$q_I > \frac{2\rho^2-\rho-1}{(1-\rho)^2} = \frac{2\rho^2-4\rho+2-3+3\rho}{(1-\rho)^2} = 2 - \frac{3}{1-\rho}$$

Since we assumed that $\rho = \frac{e_I^2}{e_O^2} \le 1$, the bound $2 - \frac{3}{1-\rho} < 0$ and the difference $\Delta > 0$ for all values of q_I.

Thus, using adaptive scheme for controlling number of pulses results in the lesser average error compared to the non-adaptive scheme which uses same number of pulses. Alternatively, it is possible to reduce the average require number of pulses without reducing the expected mean square error.

For a general case, we define the k-N Adaptive Optimal Weighted Averaging (AOWA) fusion scheme by using following algorithm:

1. Perform measurements , applying the classifier on each measurements
2. Stop measurement process when either k inliers has been detected or total number of N measurements has been reached
3. Fuse obtained estimates using OWA fusion scheme

As the results, the number of estimated fused during the each measurements varies depending on specific outcome but never exceeds N pulses in the worst case or fall below k pulses in the best case.

Depending on the quality of the classifications and prior probabilities for appearance of outliers, only on rare occasions the process requires significantly more than k pulses. In those cases, additional measurements are made for compensating an uncertainty associated with suspected outlier measurements.

The expected number of outliers given k inliers with unlimited number of measurements follows the Negative Binomial distribution with parameters q_I and k. Therefore, the expected number of outliers N_O conditioned on $N_O \le N-k$ can be easily computed using [8]. Therefore

$$E\{k+N_O \mid N_O \le N-k\} = \frac{k}{q_I}(1-\alpha)$$

Where

$\alpha = (1-q_I)\dfrac{p_{NB}(N-k)}{F_{NB}(N-k)} \cdot \dfrac{N}{k}$ and p_{NB}, F_{NB} are the probability mass function (PMF) and the cumulative distribution function (CDF) of negative binomial distribution with appropriate parameters.

Using the simulation process described in details in [5], it can be shown (Figure 6) that actual number of pulses rarely exceed the low limit of k pulses but when it does, it has significant impact on overall MSE of the estimation.

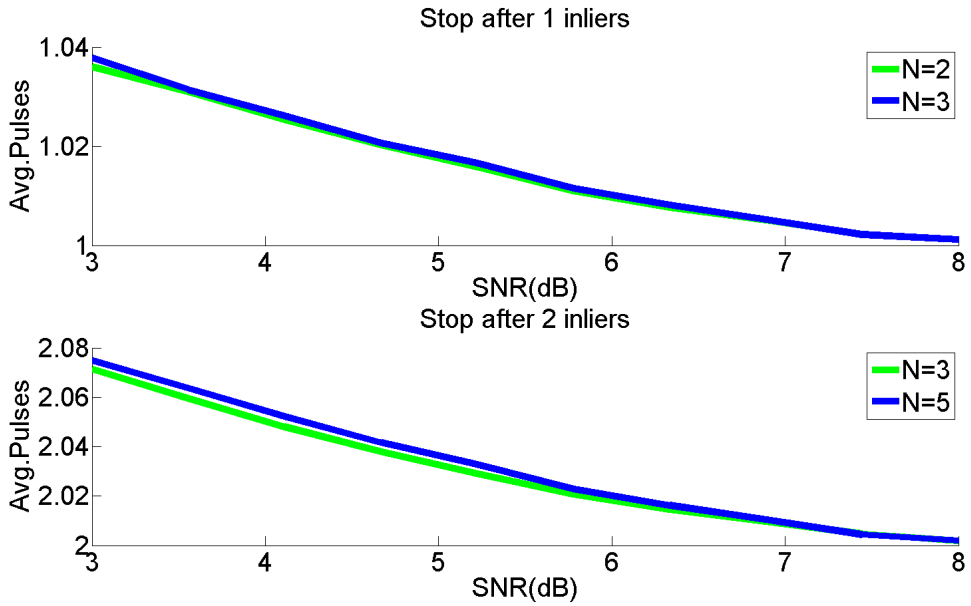

Figure 6: The average number of pulses if stopped after 1 inlier (top) or after 2 inliers (bottom). The different lines correspond to different maximum number of pulses. Only on rare occasions, additional pulses are required due to detection of outliers.

In the next section, we provide the results on the performance of AOWA method comparing to OWA and other conventional fusion methods (simple averaging and fusion by robust statistics). The simulation results show that OWA method has the best accuracy to invested energy ratio among all considered methods.

VIII. ENERGY EFFICIENCY

Since the AOWA method uses variable amount of ping and, thus, variable amount of energy, it can only be compared with other methods that uses equal amount of energy. Moreover, we only consider the settings which imply a limitation on the power of each individual pulse. Under these settings, the SNR of each pulse is within the semi-coherent region and, thus, each individual measurement is subject to the threshold. The probability of the outlier in each individual measurement is notable, but depends on the power (or, equally on SNR) of each measurement.

The most interesting question that arises under these constraints concerns with optimal usage of the energy. For instance, splitting the fixed amount of energy among larger number of pulses reduces the total MSE due to independence of noise samples in each measurements but it also significantly increase the probability of outlier measurements due to reduced power (and thus reduced SNR) of each measurement. This point is illustrated by Figure 7. Although the RMSE can be reduced by invested more energy through increasing number of pulses, increasing number of pulses while keeping the total energy causes the increase in the RMSE. This is due to the fact that independence of noise samples does not overweight the increase in the threshold effect due to lowering SNR of individual pulses. Practically, it means that if simple averaging is used for estimation with the semi-coherent region, it is better to stay with less but more powerful pulses than to employ a larger number of lower energy pulses.

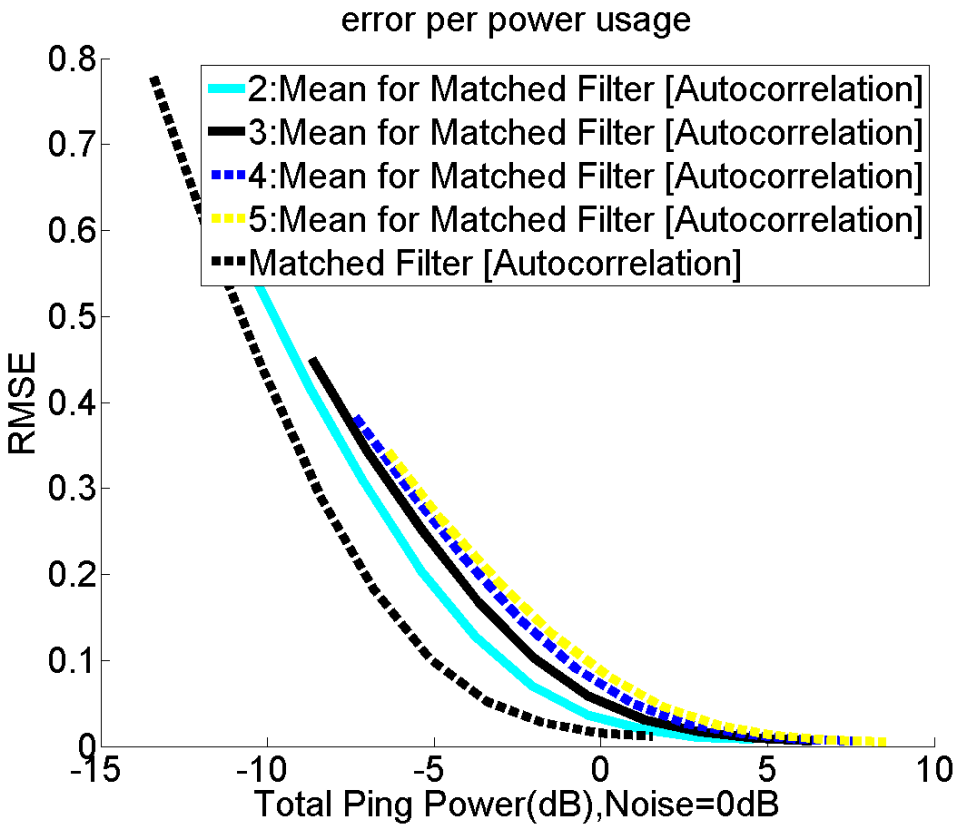

Figure 7: Although the RMSE can be reduced by invested more energy through increasing number of pulses, increasing number of pulses while keeping the total energy causes the increase in the

RMSE. This is due to the fact that independence of noise samples does not overweight the increase in the threshold effect due to lowering SNR of individual pulses.

What about the fusion by robust statistics? Does fusion by robust statistics (e.g. median) change the balance between number of pulses and the energy of an individual pulse? **Figure 8** provides a useful insight on the effect. It is clear from picture that increasing number of pulses while keeping the total energy constant does affect the RMSE of the estimation for fusion by median statistics. However, the mean fusion is still no better off than using a single pulse of combined energy.

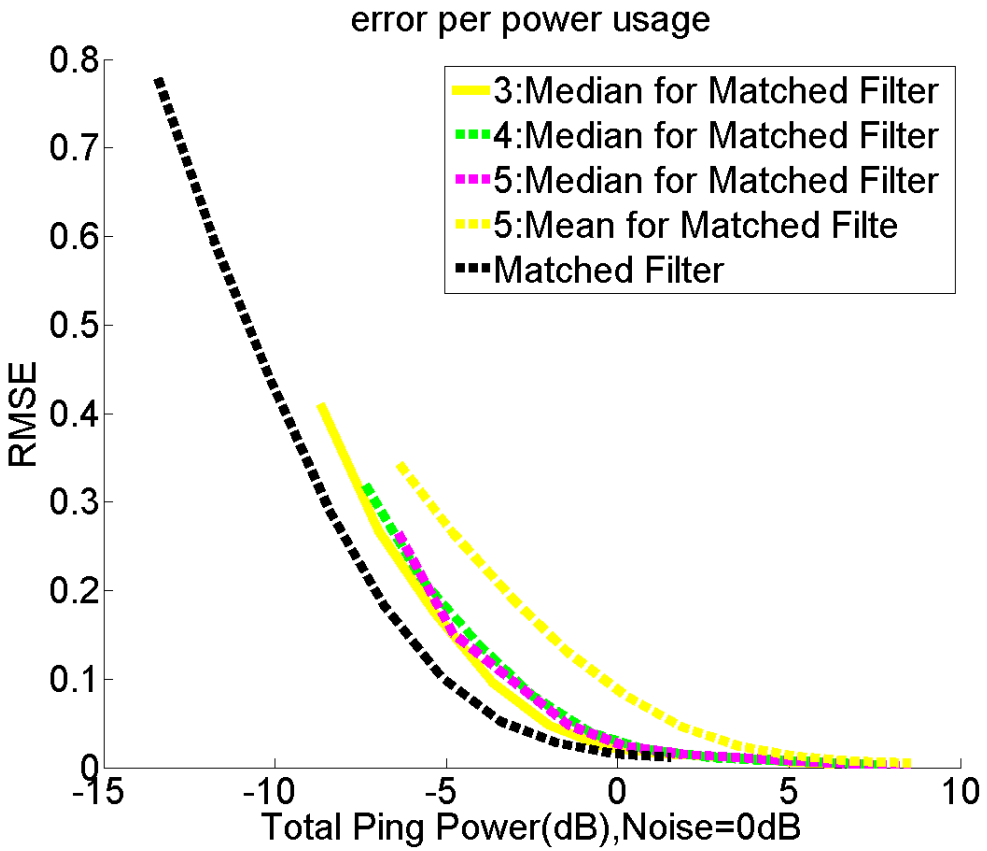

Figure 8: Increasing the number of pulses while keeping the total energy constant does affect the RMSE of the estimation for fusion by median statistics. However, the median fusion is still no better off than using a single pulse of combined energy. The fusion by simple averaging (N=5) is presented for the reference

Next, we consider the energy efficiency of non-adaptive OWA fusion method. **Figure 9** has the relevant result. For OWA the situation is reversed, that is in order to achieve higher level of the accuracy with fixed total available energy, it is better to split the energy into a number of pulses, provided the OWA fusion method is used for obtaining the final estimate.

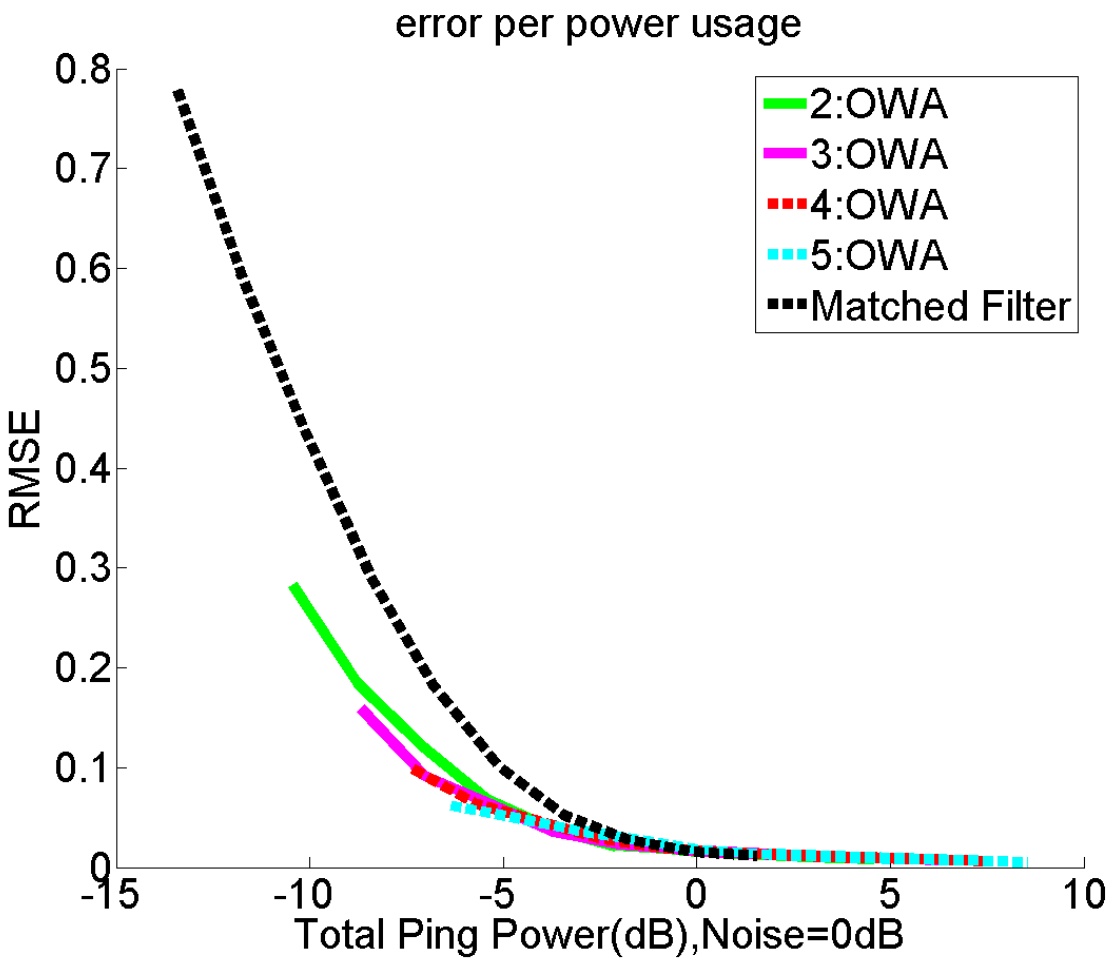

Figure 9: To achieve a higher level of the accuracy with fixed total available energy, it is better to split the energy into a number of pulses provided the OWA fusion method is used for obtaining the final estimate.

Finally, we consider the Adaptive Optimal Weighted Averaging (AOWA) method. First, let us compare the energy-efficiency of the k-N early stopping rule for different values of k and N. **Figure 10** shows RMSE lines for several values of k and N. From this data, it seems that the best strategy when employing low powered pulses is to decrease required number of inliers (k) while increasing the upper limit (N) on the total number of pulses.

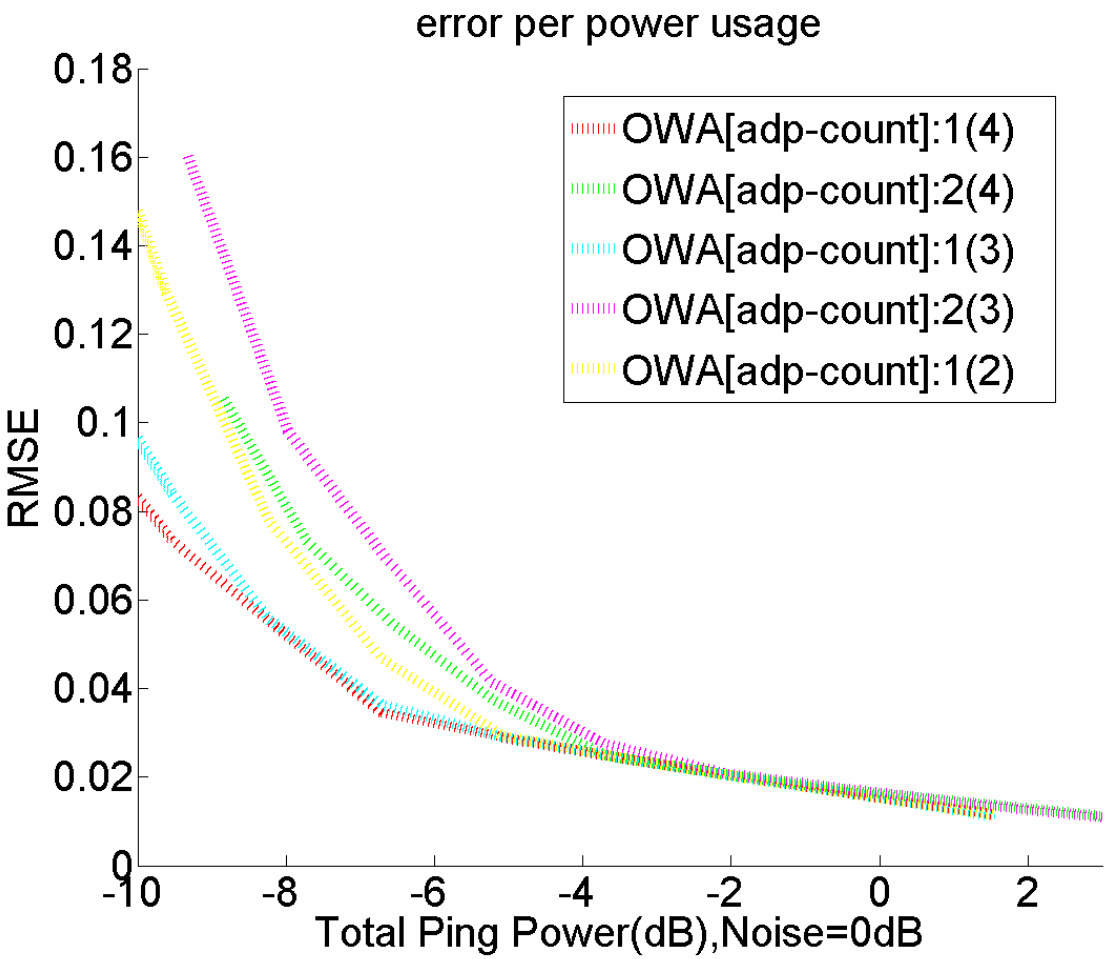

Figure 10: From this data, it seems that the best strategy when employing low powered pulses is to decrease required number of inliers (k) while increasing the upper limit (N) on the total number of pulses.

The comparison of AOWA with non-adaptive OWA counterparts is presented in **Figure 11.** The Adaptive Optimal Weighted Average method for ToF estimation in the presence of outliers outperforms all considered methods in terms of the energy efficiency.

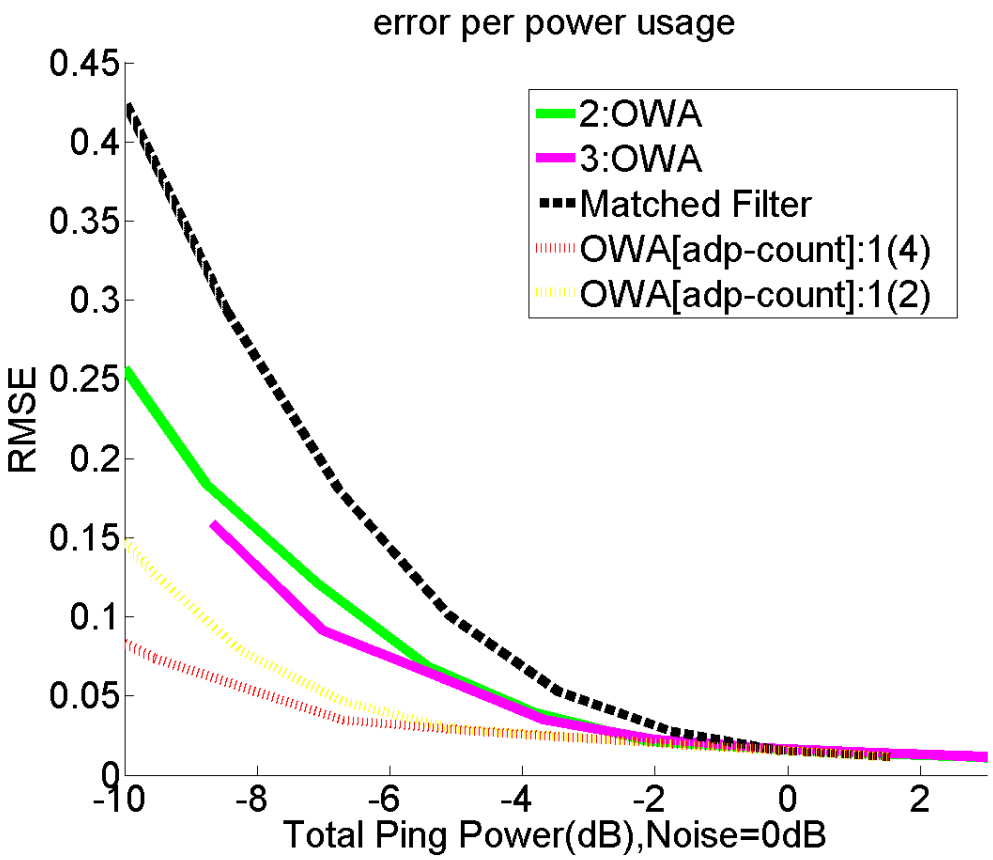

Figure 11: The Adaptive Optimal Weighted Average method for ToF estimation in the presence of outliers outperforms all considered methods in terms of the energy efficiency.

In this section, we used the computer simulation to analyze energy efficiency of different near-the-threshold ToF estimation schemes. The proposed AOWA method and earlier proposed OWA method for fusion of estimates from multiple pulses have been shown to be superior in terms of energy-efficiency compared to conventional fusion by simple averaging or even for more advanced fusion by robust statistics (median).

IX. SUMMARY AND CONCLUSIONS

In this work, we have introduced a method for improving near-threshold ToF estimation. Since the threshold effect in this problem emerges due to the multimodal shape of the ensemble-average likelihood function, it is possible to reduce the threshold effect through usage of a weak classifier. After the classifier has been trained, its output could be used for computing and, subsequently, subtracting from the MLE the expected bias occurring due to these outliers. The second contribution of this work is the introduction of the phase shifted unmatched filters as a means to create a collection of biased estimators. These estimators, were used to generate a feature vector that characterizes the maxima of likelihood function. The simulation results showed the combined effect of these two approaches on near-threshold MSE of Time of Flight MLE.

We also have introduced a method for combining individual estimates into a single robust ToF estimate in the presence of outliers. Using labels supplied by the classifier, we have computed the optimal weights to be in the fusion by the weighted averaging. The weights have been assigned in a manner that assures smaller weights to less certain estimates.

Finally, we have analyzed the whole estimation process in terms of the energy utilization. We have shown that described OWA fusion method is different from the conventional simple averaging and fusion by median statistics from energy utilization perspective. The described OWA method allows increasing the estimation accuracy by splitting available energy into multiple pulses. Moreover, we have proposed a method for adaptive control over the estimation process. The described Adaptive OWA method allows an additional improvement in estimation accuracy while using same amount of energy.

All proposed methods can be employed during post-processing phase of the measurements and does not require altering the shape of the source pulse. Therefore, these methods can be easily applied in practical applications for improving the accuracy and energy-efficiency of Sonar and other remote sensing applications.

The methods developed under this project improve ToF estimation accuracy and energy consumptions for ToF estimation under very low Signal-To-Noise Ratio (SNR). Operating under low SNR is frequently necessary in many military applications as it might require keeping the power of source pulses low for staying undetected by the adversary or when the exploration is performed by mobile robots with autonomous and, thus, limited power source (battery). The proposed methods reduce the energy requirements of the ToF estimation, while significantly improving the accuracy of ToF measurements under low SNR. These proposed methods, thus bring significant benefits for a wide class of practical military applications.

X. REFERENCES

[1] Alpayadin, E., “Introduction to Machine Learning”, MIT Press 2004

[2] A. Apartsin, LN. Cooper, N. Intrator ,“*Biosonar-inspired source localization in low SNR*”, Proceedings of BIOSIGNALS’2011, pp.399-404

[3] A. Apartsin, LN. Cooper, N. Intrator, “*SNR-dependant filtering for Time of Arrival Estimation in High Noise*”, Proceedings of IEEE Machine Learning for Signal Processing(MLSP) 2010, pp427-431

[4] A.Apartsin,LN.Cooper,N.Intrator,“Semi-coherent Time of Arrival Estimation Using Regression”, Journal of the Acoustic Society Of America(JASA),132(2), 2012

[5] E.W. Barankin, “*Locally best unbiased estimates*” The Annals of Mathematical Statistics (20), 1949, pp477-501

[6] B.M Sadler, R.J. Kozick, “*A Survey of Time Delay Estimation Performance Bounds”*, Fourth IEEE Workshop on Sensor Array and Multichannel Processing, 2006, pp. 282-288

[7] S. M. Kay, "*Fundamental of Statistical Signal Processing*", Prentice Hall, New Jersey, 1993, 595 pages

[8] S. S. Keerthi, C.J. Lin, "*Asymptotic behaviors of support vector machines with Gaussian kernel*" Neural Computation, 15(7), 2003

[9] T. Kimchi, M. Reshef**,** J. Terkel, "*Evidence for the use of reflected self-generated seismic waves for spatial orientation in a blind subterranean mammal*", Journal of Experimental Biology 2005, pp647-659

[10] R.N. McDonough, A. D. Whalen, "*Detection of Signals in Noise*", Academic Press, San Diego ,1995.495 pages

[11] N. Neretti, N. Intrator, and L. N Cooper, "*Adaptive pulse optimization for improved sonar range accuracy*." IEEE Signal Processing Letters 11(4), 2004, pp. 409-412

[12] N. Neretti, M.I. Sanderson, J.A Simmons, and N Intrator, "*Time-frequency computational model for echo-delay resolution in sonar images of the big brown bat, Eptesicus fuscus*", Journal of the Acoustic Society of America 113(4), 2003, pp2137-2145

[13] D.R. Pauluzzi, N.C. Beaulieu, "*A comparison of SNR estimation techniques for the AWGN channel*", IEEE Transaction on Communications 2000, pp1681-1691

[14] A. Pinkus, J. Tabrikian, *"Barankin bound for range and Doppler estimation using orthogonal signal transmission",* in Proceedings of IEEE Radar Conference, 2006, pp. 94–99

[15] Rudin, W.,"Real and Complex Analysis", McGraw-Hill 1986

[16] M. I. Sanderson, N. Neretti, N. Intrator and J. A. Simmons, *"Evaluation*." Journal of the Acoustical Society of America 114(3), pp. 1648-59, September 2003

[17] J.A. Simmons and R. A. Stein, "A*coustic imaging in bat sonar: Echolocation signals and the evolution of echolocation.*" Journal of Comparative Physiology, 135(1), 1980, pp61-68

[18] M.I. Skolnik, "*Introduction to Radar Systems*", McGraw-Hill, New York, 1962, 772 pages

[19] GP. Succi, G. Prado, R. Gampert and TK. Pedersen, "*Problems in Seismic detection and Tracking*", Proceedings of SPIE 2000, pp165-173

[20] Harry L. Van Trees, "*Detection, Estimation and Modulation Theory*", John Wiley & Sons, Inc., New York, 2004, 720 pages

[21] L. R. Varshney and D. Thomas, "*Side lobe reduction for matched filter range processing*", Proceedings of IEEE Radar Conference, 2003, pp446-451

[22] P.D. Wasserman, "*Advanced Methods in Neural Computing"*, New York, Van Nostrand Reinhold, 1993, pp. 155–61

[23] P. Woodward, "*Probability and Information Theory, with Applications to Radar*", McGraw-Hill, New York, 1953,140 pages

[24] L. Yu, N. Neretti, N. Intrator, "*Multiple ping sonar accuracy improvement using robust motion estimation and ping fusion*", JASA 2006, pp2106-2113

[25] A. Zeira and P.M. Schultheiss, "*Realizable lower bounds for time delay estimation-Part II: Threshold phenomena*" , IEEE Trans. Signal Processing,1994, pp1001-1007

[26] Elfes, A., "*Sonar-based real-world mapping and navigation*", IEEE Journal of Robotics and Automation, 1987, pp249-265

[27] Chapman, C., "*Fundamentals of Seismic Wave Propagation*", Cambridge University Press, 2004

PUBLICATION EMANATING FROM THE CURRENT RESEARCH

Conference papers:

[1] A. Apartsin, LN. Cooper, N. Intrator ,"*Biosonar-inspired source localization in low SNR*", Proceedings of BIOSIGNALS'2011, pp.399-404

[2] A. Apartsin, LN. Cooper, N. Intrator, "*SNR-dependant filtering for Time of Arrival Estimation in High Noise*", Proceedings of IEEE Machine Learning for Signal Processing(MLSP) 2010, pp427-431

Journal papers:

[1] A.Apartsin,LN.Cooper,N.Intrator,"Semi-coherent Time of Arrival Estimation Using Regression", Journal of the Acoustic Society Of America(JASA),132(2), 2012

[2] A.Apartsin,LN.Cooper,N.Intrator,"Time of Flight Estimation in the Presence of Outliers Part I-Single Echo Processing", 2012 (submitted)

[3] A.Apartsin,LN.Cooper,N.Intrator,"Time of Flight Estimation in the Presence of Outliers Part II-Multiple Echo Processing", 2012 (submitted)

[4] A.Apartsin,LN.Cooper,N.Intrator,"Time of Flight Estimation in the Presence of Outliers Part III-Energy Efficiency", 2012 (in preparations)